DOCUMENT RESUME

ED 482 934 TM 035 326

AUTHOR	Glass, Gene V., Ed.
TITLE	Education Policy Analysis Archives, 2003.
ISSN	ISSN-1068-2341
PUB DATE	2003-00-00
NOTE	658p.; "Education Policy Analysis Archives" is an electronic-only journal (see http://olam.edu.asu.edu/epaa) covered on an article by article basis in "Current Index to Journals in Education" (CIJE). For the 37 English language articles in this volume, see TM 525 625-661.
AVAILABLE FROM	For full text: http://olam.ed.asu.edu/epaa.
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Education Policy Analysis Archives; v11 n1-39 2003
EDRS PRICE	EDRS Price MF03/PC27 Plus Postage.
DESCRIPTORS	*Academic Achievement; Case Studies; Charter Schools; *Educational Change; *Educational Research; *Elementary Secondary Education; *Higher Education; Professional Development; School Choice; State Programs; *Testing Programs

ABSTRACT

This document consists of articles 1 through 39 published in the electronic journal "Education Policy Analysis Archives" for the year 2003. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

VOLUME 11 - 2003

ISSN 1068-2341

ED 482 934

EDUCATION POLICY ANALYSIS ARCHIVES

(Articles 1 thru 39)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

G. Glass

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

TM035326

editor:
Gene V. Glass
Arizona State University

BEST COPY AVAILABLE

EDUCATION POLICY ANALYSIS ARCHIVES

Volume 11, 2003

Article **Page No.**

1 Dorn: High-Stakes Testing and the History of Graduation 18

2 Willinsky: Policymakers' Online Use of Academic Research 14

3 Mintrop: The Limits of Sanctions in Low-Performing Schools 17

4 Milanowski: Varieties of Knowledge and Skill-Based Pay Design 23

5 Dowson et al.: Continuing Education Reform in Hong Kong 17

6 Cervini: Diferencias de resultados cognitivos y no-cognitivos entre estudiantes de escuelas públicas y privadas en la educación secundaria de Argentina 20

7 Avery, Beach & Coler: Impact of Minnesota's "Profile of Learning" 22

8 Buchanan & Fox: Case Studies of Emerging Ethnocentric Charter Schools in Hawai'i 14

9 Meisels et al.: Impact of Instructional Assessment on Elementary Children's Achievement 11

10 Linton & Kester: Exploring the Achievement Gap Between White & Minority Students in Texas 13

11 Hawes & Donoso: Análisis del concepto de "la docencia" en profesores universitarios 32

12 Scribner: Teacher learning in context: The special case of rural high school teachers 14

13 Fowler: Massachusetts Signing Bonus Program for New Teachers 15

14 Hu: Educational Aspirations and Postsecondary Access and Choice 9

15 Camilli, Vargas and Yurecko: Teaching Children to Read : The Fragile Link Between Science & Federal Education Policy 33

16 McMeekin: Networks of schools 13

17 Gorostiaga, Acedo & Xifra: Secondary Education in Argentina during the 1990s 17

18 Orlich: Longitudinal Effect of the Washington Assessment of Student Learning (WASL) on Student Achievement 9

19 Wenglinsky: Using Large-Scale Research to Gauge the Impact of Instructional Practices on Student Reading Comprehension: An Exploratory Study 12

20 Zvoch & Stevens: A Multilevel, Longitudinal Analysis of Middle School Math and Language Achievement 14

21 Ruano: Some Salient Features of Guatemala's Public and Private School Sectors 17

22 Mabry, Poole, Redmond & Schultz: Local Impact of State Testing in Southwest Washington 22

23 Murphy, DeArmond & Guin: A National Crisis or Localized Problems? Getting Perspective on the Scope and Scale of the Teacher Shortage 12

24 Rosenshine: High-stakes Testing: Another Analysis 5

25 Amrein-Beardsley & Berliner: Re-analysis of NAEP Math and Reading Scores in States with and without High-stakes Tests: Response to Rosenshine 11

26 Stritikus & Garcia: The Role of Theory and Policy in the Educational Treatment of Language Minority Students: Competitive Structures in California 19

27 Weiner: Research or "Cheerleading"? Scholarship on Community School District 2, New York City 13

28 Resnick: Reforms, Research and Variability: A Reply to Lois Weiner 5

29 Goodman: Comments on Weiner, Resnick and Scientific Debate 3

30 Ferrero: Embracing Pedagogical Pluralism: An Educator's Case for (at Least Public) School Choice 13

31 Linn: Performance Standards—Utility for Different Uses of Assessments 13

32 Frankenberg & Lee: Charter Schools and Race: A Lost Opportunity for Integrated Education 29

33 Darling-Hammond & Sykes: Wanted, A National Teacher Supply Policy for Education: The Right Way to Meet The "Highly Qualified Teacher" Challenge 34

34 Mathison & Freeman: Constraining Elementary Teachers' Work: Dilemmas and Paradoxes Created by State Mandated Testing 15

35 Gallucci: Communities of Practice and the Mediation of Teachers' Responses to Standards-based Reform 19

36 Gándara,P., Rumberger, R., Maxwell-Jolly, J. & Callahan, R.: English Learners in California Schools: Unequal resources—Unequal outcomes 34

37 Halverson: Systems of Practice: How Leaders Use Artifacts to Create Professional Community in Schools 22

38 Warschauer: The Allures and Illusions of Modernity: Technology and Educational Reform in Egypt 15

39 Hill & Guin: Baselines for Assessment of Choice Programs 20

This article has been retrieved 8820 times since January 1, 2003

Education Policy Analysis Archives

Volume 11 Number 1 **January 1, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the **EDUCATION POLICY ANALYSIS ARCHIVES**. Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

High-Stakes Testing and the History of Graduation

Sherman Dorn
University of South Florida

Citation: Dorn, S. (2003, January 1). High-Stakes testing and the history of graduation, *Education Policy Analysis Archives, 11*(1). Retrieved [date] from http://epaa.asu.edu/epaa/v11n1/.

Abstract
An historical perspective on high-stakes testing suggests that tests required for high school graduation will have mixed results for the putative value of high school diplomas: (1) graduation requirements are likely to have indirect as well as direct effects on the likelihood of graduating; (2) the proliferation of different exit documents may dilute efforts to improve the education of *all* students; and (3) graduation requirements remain unlikely to disentangle the general cultural confusion in the U.S. about the purpose of secondary education and a high school diploma, especially confusion about whether the educational, exchange, or other value of a diploma is most important.

Introduction

With the ascension of George W. Bush to the presidency, the scripting of national education policy debates has become more predictable. In the first week of his term, Bush proposed the same set of policies he had advocated or tacitly supported while governor of Texas, including more reliance on standardized testing to judge schools and government support for private (including parochial) schooling. During the 2000 campaign, he touted Texas's reform record in education as evidence of his competence in education policy. However, the details of education policies focusing on testing have become murkier, not clearer, in individual states. One issue in which debate over high-stakes testing has not settled down has been in the area of graduation testing. In the past two years, the failure of a federal court case against the Texas graduation test requirements, the publication of one expert witness's research on Texas, and two RAND Corporation studies that analyzed test scores from Texas have increased academic and civil-rights scrutiny of high-stakes testing as a gatekeeper for high school graduation (*GI Forum v. Texas Education Agency*, 2000; Grissmer, Flanagan, Kawata, & Williamson, 2000; Haney, 2000; Klein, Hamilton, McCaffrey, & Stecher, 2000; National Research Council, 2001). Texas's state policies have passed judicial—but perhaps not broader—tests. Various other state education agencies have retreated from fast implementation schedules for higher-stakes tests, especially as high failure rates have attracted attention. Lisa Keegan, Arizona's former superintendent of schools, was notable in shifting her ground several times as parents questioned the validity of test results, especially their weak connection to what schools may have taught the tenth graders subject to the testing (Kossan, 2000) and to skills used by the workforce (Glass & Edholm, 2002).

The academic and legal debate over graduation tests may well seem like *deja vu* to those who remember implementation of minimum competency tests in the 1970s and early 1980s (*Debra P. v. Turlington*, 1981, 1983; Linn, 2000; Madaus, 1983; McDill, Natriello, & Pallas, 1985, 1986; Pullin, 1981; Salganik, 1985; Turlington, 1985). The "script," so to speak, is familiar. As was the case with minimum competency tests, advocates of graduation tests today argue that one must test potential students to ensure that a high school diploma "means something." As in the 1980s, others are concerned that graduation tests are a substantial barrier to education that encourages dropping out, with differential impacts on poor and minority children. As in the 1980s, whether existing data can be used to answer the empirical questions about the impact of graduation tests is doubtful. And, as in the 1980s, one major federal court case left the test requirement in a large state essentially intact.

To an historian, the debate over high-stakes graduation tests is a recent phenomenon that can only take place because the majority of teenagers graduate. The contemporary concern with the consequences of high-stakes testing on graduation could not have existed a century ago. Few students attended high school, graduation was not expected of teenagers, and although American utopian writer Edward Bellamy had coined the expression "to drop out" of school, it would not become the dominant form of describing those who leave school until the 1960s (Dorn, 1993; Dorn & Johanningmeier, 1999). In the past forty years—but not before—educational researchers and administrators have written about dropping out as a serious social problem. In the past few years, some have renewed debate about the equity concerns with high school graduation, specifically with high-stakes graduation tests (e.g., Haney, 2000; National Research Council, 2001). What is notable is the apparent contradiction between the putative national goal of increasing the likelihood of graduation, on the one hand, and the increasing high stakes of standardized testing as the mechanism used for both student and school accountability in many states. This paradox has its roots in the history of high schools in the twentieth century and changing expectations for teenagers.

Three topics are important in tying together the history of high school graduation and contemporary high-stakes testing policies. First is the statistical evidence about broad trends in high school attendance and graduation, insofar as sociologists, demographers, and historians have gathered them. The likelihood of teenagers graduating from regular high school programs increased dramatically in the first two thirds of the century but has stabilized since. Second is a discussion about the proliferation of diplomas. Apart from the debate over the value of the GED, there has been little discussion not only on the growing differentiation of diplomas but also on how that differentiation fits into the larger history of secondary and higher education in the U.S.. Schools are, for better or worse, more prepared to make additional distinctions in programs and diplomas than they are to prepare most children for a single credential. Last is an analysis of how public policy discussion of dropping out reflects the relatively new expectation of graduation. That new norm is the latest form in which we place our expectations for schooling more broadly. Our assumption that schools should solve social problems has colored popular images of dropouts, continues to shape official dropout policies, and explains why the debate about dropping out has, historically, omitted key issues.

The history suggests several general conclusions about the relationship between high-stakes testing and graduation patterns. First is how dropping out may be larger as an indirect than as a direct effect of high-stakes testing. A complex web of influences shapes entrances into and exits out of school as well as engagement with learning. One of these factors known to historians of education is the reciprocal relationship between labor-market participation and school attendance, and it is this type of push-and-pull relationship that high-stakes testing may affect. If barriers to grade promotion result in a large group of students who are clearly unable to graduate from high school by 19 years of age (when the majority of their age peers have already graduated and left high school), other opportunities (primarily work) will beckon strongly.

The second reasonable conclusion from this history is that this proliferation in diplomas, whether at the secondary or tertiary level, may mask continuing inequalities in educational opportunity. There seem to be many opportunities to earn degrees in various ways, at various times in life. And, because of the high proportion of the population attending college *at some point* in life, one might claim that high school graduation is merely an intermediate step in formal schooling. It is, one might say, just on the way to something that has a larger payoff, in the same way that completing eighth grade has ceased to be meaningful for most students' long-term future. However, because the high school degree is still a prerequisite for more advanced formal schooling, one must still pay attention to it as a gateway for schooling in adulthood as well as direct labor-market consequences of raising barriers to high school graduation.

Third, the history of dropping out, both as demography and as a public policy concern, suggests the difficulty of disentangling key issues. Can the public intelligently discuss the potential values of a high-school diploma when North American society has conflated them for almost forty years? We have inherited a legacy of viewing dropping out primarily as a problem of human capital, ignoring the issue of credentialism; a problem of future dependency, diminishing the equity problem; and a problem of individual psychology, effacing the broader social factors which Rumberger (1987, 1995), among others, has described over the years; and a problem whose solution is less important than establishing the "worth" of a credential, overlooking how the two are intimately bound together. Open public debate should honestly face the dilemmas of a popular education in a high-stakes testing environment, even if the barriers to such debate are massive (Cremin, 1989).

In part, we have difficulty distinguishing the relative importance of the education credential as an exchangeable result (for more schooling or better work), on the one hand, and the knowledge and skills a student can take from education to use directly, on the other. We have inherited, from the twentieth-century history of educational development and debate about schooling, an assumption that the value of a diploma is or should be synonymous with the knowledge and skills a student presumably gains from schooling. With the first concerns about dropping out as a social problem in the 1960s came statistical evidence about how much more graduates earn than dropouts (Dorn, 1996). The existence of alternative explanations—the potential for credentials to sort labor queues, for example, or preexisting advantages that correlated both with educational attainment and also adult labor-market success—remained obscure or invisible. Over the same century, institutions of secondary and higher education have invented more and more ways to differentiate education, either implicitly through tracking and the creation of selective institutions or explicitly through different degrees. The last fifteen years have witnessed an expansion in high-school exit documents, and whether this is a unique moment in the history of the diplomas as such, it fits into the historical pattern of differentiation to solve the pressures on schools as institutions. High schools and colleges have tried, over the past century, both to grant more access and also to gain the institutional rewards of restricting access to the most valued credentials. Their solution—differentiation—suggests how strongly our society pushes schools to achieve *both* ends, perhaps at the cost of clarity about the fundamental purposes of schooling.

Numbers

Researchers across social science disciplines have come to very similar conclusions about the pattern of earning high school diplomas over the past century. Though some of the details vary among the authors, there is general agreement about the following broad trends:

The Twentieth Century has witnessed...

- increased high school graduation rates over the first two-thirds of the twentieth century, and stabilization since;
- decreased racial/ethnic and gender gaps in graduation during general increasing graduation, but persisting racial gaps since;
- growth in alternative credentials in the past several decades, and different rates in earning alternative credentials by various population groups; and
- persistent socioeconomic influences on graduation.

The crucial questions for determining the fairness of testing as a gatekeeper for academic diplomas revolve around why these patterns have developed and the theoretical potential for high-stakes tests to magnify continuing inequalities of educational opportunities. In particular, the reciprocal relationship between labor-market participation and school attendance is of particular note for how work may become more and more attractive as teenagers think of themselves as further away from high school graduation.

Broad Trends

At the beginning of the twentieth century, fewer than one of every ten adolescents graduated from high school. Today, roughly seven of every ten teens can expect to earn a diploma through a regular high school program (Goldin, 1999; Snyder & Hoffman, 2002). The increase, occurring in the first seven decades of the twentieth century, represents a dramatic change in the educational attainment of children in the United States. While elementary schooling was widespread in the nineteenth century, attendance was sporadic, and a minority attended (let alone graduated from) secondary schools. Today, formal schooling dominates children's daily lives as a part of growing up, and as a part of dominant beliefs about succeeding as adults. First, most children attend formal schools regularly, and as a result, schools circumscribe the lives and concerns of most families who schedule vacations with the school calendar in hand and who plan their daily lives around school schedules. Second, children attend school through most of their childhood. In contrast to the patterns of one hundred years ago, when children typically left formal schooling in their early teens, most children today attend school until they are legal adults. Third, schooling *as a route to economic and social success* has become part of the American belief in the existence of inequality without social classes (DeMott, 1990; Ossowski, 1963). Most children would agree, as would most adults, that a good education is a requirement for a good job. Few would have agreed with such a statement in 1890, even though elementary schooling had become an accepted, politically popular part of childhood in the nineteenth century (Katznelson & Weir, 1985). The expansion of secondary education has coincided with a growing rationale of schooling as a way to improve students' future job prospects (Kantor, 1983).

Figure 1 (which overlays data from three different sources by birth cohort (Note 1) illustrates both the broader trend and the two significant exceptions to this pattern of increasing graduation: the decrease in graduation at the end of World War II and the stability in graduation from regular high school programs over the past several decades. The ratio of high school graduates to 17-year-olds rose to 0.70 for the cohort born in 1943 (and graduating in 1960 or shortly afterwards), and has remained approximately at or above 0.70 since. But the ratio had **declined** from 0.51 for those born in 1925 to 0.43 for those born in 1927 and 0.47 for the 1929 cohort before returning to 0.53 for the 1931 cohort—corresponding to a dip in graduation between 1942 and 1948, the period when those cohorts would have reached 17 years of age (Goldin, 1999, p. 63). This pattern mirrored decreases in high school attendance during World War II (Snyder, 1993). Because the drop included females as well as males (though it was more precipitous for males), military service cannot entirely explain the temporary shift; teens left school primarily to work during the war (Goldin, 1998). More recently, the likelihood of graduating from regular high school programs—as opposed to alternative credentials such as through the General Educational Development (GED) test—has remained fairly stable (with a slight decrease) over the past three decades or more, in contrast to the prior dramatic increase. The self-reported graduation proportions (reported by the Census Bureau) are higher than ratio of diplomas to 17-year-olds, a series which has decreased slightly since the cohorts born in the mid-1950s.

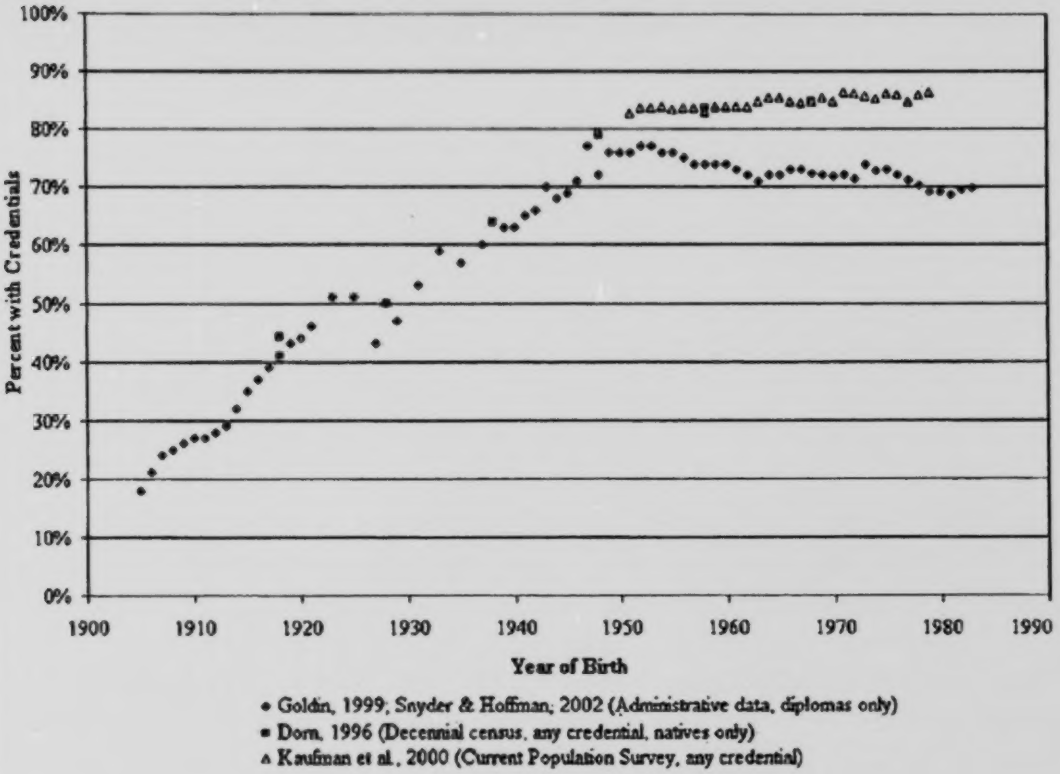


Figure 1. Twentieth Century U.S. Graduation Rate By Birth Cohort and Data Source

(The data for Figure 1 may be downloaded in the form of an Excel worksheet.))

During the increase in graduation before those born in the 1950s, three types of gap in the likelihood of graduating shrank dramatically: by race, geographic region, and gender. According to Dorn (1996), the white-nonwhite difference in having high school credentials by 20-24 year olds shrank from a 34 percent absolute gap in 1940 to a 13 percent gap in 1980. Other research (Jaynes & Williams 1989; Kaufman, Alt, & Chapman, 2001) confirms the shrinking racial gap in high school graduation. Some of that difference early in the century came from the concentration of African Americans in the South before World War II. Goldin's data (1999) show that while whites in the Southern census divisions had better graduation rates than the regional population as a whole, Southern whites were still less likely to graduate than those from other regions through the mid-1950s. Multivariate analysis shows that both region and race became less important, in themselves, in determining who graduated by the last third of the century, but both had independent associations with the odds of graduating earlier (e.g., Dorn, 1996; Featherman & Hauser, 1976). The relative influence of gender has also dwindled. In 1910, 93,000 females graduated, as opposed to 64,000 males, almost a 3:2 ratio. As high school graduation became far more common, boys began to catch up with girls in basic educational attainment. In 1970, 1.46 million females graduated, in contrast to 1.43 million males. In 1991, for the first time since the federal government began collecting records on high school graduation, more males graduated than females (by an estimated 3,000 out of more than 1.2 million for each sex) (Goldin, 1999, pp. 63-64).

A racial gap persists in regular high school graduation, though, in terms of absolute percentages (Dorn, 1996; Kaufman et al., 2001). This gap parallels, to some extent, concerns about declines in the fortunes of African Americans in higher education in the 1980s (Jaynes & Williams, 1989). Given the greater likelihood of minority children being poor, perhaps one can explain this continuing gap as an indirect effect of persistent socioeconomic or social class effects on educational attainment. Some evidence, however, is more disturbing. While the influence of being African American on the odds of graduation by age 18 had disappeared by 1970 in my multivariate analysis of decennial censuses, the association reappeared in 1980 and 1990 for graduation by age 19 (Dorn, 1996, p. 20). I hypothesized that race had disappeared as a factor in the chances of *early* graduation after "socioeconomics/social class" is factored out, but that African Americans are still less likely, even in a multivariate analysis, to be regular high school graduates by the end of the ages that most youth attend high school. In other words, the proportion of African Americans who are slightly precocious academically (who would graduate before their eighteenth birthday), given their economic circumstances, is probably as high as the general population. Others, who may be less resilient, might be far more vulnerable to the combined effects of poverty and racism. Recent reports of racial differences in who earns regular diplomas as opposed to alternative credentials (primarily the GED) suggest some confirmation of this suggestion (Kaufman et al., 2000). Of those who earn some high school credential, white students are more likely than the general population to earn high school diplomas, while Latino/a and African American students are more likely to have alternative credentials.

These alternative credentials, primarily the GED, have grown dramatically in the last three decades both in terms of programs and also in terms of those who receive a diploma alternative. After 1970, when amendments to the Adult Education Act allowed federally-sponsored adult education activities to enroll minors and to lead to an alternative credential, growth snowballed to 294,000 GED recipients in 1974, 489,000 in 1981, and 516,000 in 1999. GED recipients represented 9 percent of young adults with high school credentials in the late 1990s (National Center for Education Statistics, 2001). Young white or Asian adults with high school credentials are more likely to have regular diplomas than young Latino/a or African American adults (Kaufman et al., 2000, p. 19). One can, of course, interpret this information in at least two ways. Maybe alternative credential programs have provided an avenue to an essential credential for disadvantaged teens that otherwise would not exist. Or maybe the growth in alternative credential programs has provided a safety valve for high schools, enabling them to slough off responsibility for poor students. The answer depends heavily on the relative value of alternative credential programs, a matter of substantial dispute (Boesel, Alsalam, & Smith, 1996; Cameron & Heckman, 1993; Murnane, Willett, & Tyler, 1999).

What is undisputed is the persistent association between socioeconomic or class status and the chances of attending and graduating from high school. The data sources vary from historical studies of local or state school systems (e.g., Kaestle & Vinovskis, 1980; Katz, Doucet, & Stern, 1982; Perlmann, 1988) to successive cross-sections of decennial censuses (Dorn, 1996) and other historical and contemporary studies (e.g., Ekstrom, Guertz, Pollack, & Rock, 1987; Featherman & Hauser, 1976; Mare, 1980, 1981; Natriello, Pallas, & McDill, 1987; Walters & Briggs, 1993). What is less certain are the influences of more family-specific factors. Historical sources generally agree with contemporary sources that having a parent or guardian with high educational attainment assists students in graduating, and that having a large family is a disadvantage (e.g., Dorn, 1996). However, those data about family structure are generally available only for the past. Many of the potential factors considered in contemporary analyses, especially psychological and school-specific factors (e.g., Rumberger, 1987, 1995; Rumberger & Larson, 1998; Rumbert & Scott, 2000), simply do not have appropriate sources providing data across a wide span of time.

What is crucial to draw from this broad history is that the debate over high-stakes testing assumes that most students graduate, a fact that developed over the twentieth century. In the past wave of implementing tests (minimum competency tests, starting in the 1970s), most appeared to continue graduating after several chances. Unfortunately, most of the data sources that might resolve this historical question about minimum competency tests are insufficient. To my knowledge, no state-level officials have intentionally proposed high-stakes testing as a means to *reduce* the likelihood of graduating, though statements such as former President Clinton's (1999) about wanting to ensure that a high school graduate can read his or her diploma are quite common. The unexamined assumption is that, given a clear set of expectations, students and schools will put forth sufficient effort to meet them. Was variation in effort responsible for the patterns of graduation in the twentieth century?

Why Growth and then Stability?

The most persuasive explanations for the patterns in graduation are theories of reciprocal movements between labor markets and schools, incentives for school systems to increase attainment to some limit, and high school custodialism. These are certainly not mutually exclusive explanations, and they each present important insights into the twentieth-century dynamics of high school graduation. Theories that focus on changes in teen labor-markets or in the accepted mission of high schools, in particular, should be a warning to those who think that simple incentives in a school system will result in clearly predictable outcomes. Complex interactions among labor markets, public debate, and social beliefs about the value of education all are likely to interfere with any policy designed with a simple model of how to improve student achievement.

Labor market-school reciprocity

The labor-market demand explanations of graduation trends focus both on the mutual exclusivity of full-time work and schooling, on the one hand, and the expectations that education can improve job prospects, on the other. As the nineteenth century turned into the twentieth, teens could find employment in a variety of fields nationwide, and someone working full-time could not also be in school. At the same time, many teens (especially boys) could see no necessary, obvious reward to continued schooling beyond age 13 or 14. Most jobs did not require, either nominally or in fact, the skills that one would develop in high school. What changed over the next sixty years was both the widespread exclusion of teenagers from full-time work and also the growing belief that an education was the ticket to success in the United States. At first in industrial occupations and, later, in agriculture, both laws and alternative opportunities discouraged employers from hiring or keeping teens as full-time laborers (Osterman, 1980).

At the same time, schools began to have a concrete reward attached to staying *in* school. The new clerical and white-collar occupations in the twentieth century were different from clerical work in the nineteenth, which was a putative stepping-stone to ownership of small companies or factories. The new division of work in the industrial era was part of a larger reorganization of businesses that concentrated the manipulation of information as well as people and things. It was the era of industrial monopoly and also of female typing pools (Chandler, 1977; Davies, 1982), and the conjunction was neither a coincidence nor inconsequential for education. With one or two years of additional attendance, especially in commercial courses, a teenager could markedly improve her or his opportunities for employment (Cohen, 1992; Kantor, 1982). One must be careful here not to overgeneralize. During the nineteenth century, the popularity of many urban high schools depended on the relative credential value of attendance (Labaree, 1988). However, the late nineteenth and early twentieth centuries witnessed the development of a new type of work, marked both by its non-manual nature and also by its wage status. With an incentive to improve one's marketable skills, teenagers in the twentieth century were more likely to attend—and thus graduate from—high schools (Goldin, 1998).

The labor-market explanation fits well with the three important features of graduation in the twentieth century. During most of the first seven decades of the century, full-time employment opportunities for teens declined, and both attendance and graduation increased at the same time. During World War II, both military service and a tight labor market encouraged teens to leave high school, temporarily reversing a decades long trend. More recently, over the past three decades, as teens have become more involved in part-time work, the mutually exclusive nature of work and school has dwindled (Greenberger & Steinberg, 1986). Perhaps some teens are less willing to believe in the long-term rewards of high school when they can earn what seems to them to be good money immediately (also see Fine, 1987). There is still a real value of diplomas in the labor market, which is the incentive for dropouts to earn GED (or other alternative) diplomas, but the relationship between the labor market and schools revolves as much around perceived rewards to and opportunity costs of schooling as the reality. The labor-market explanation is plausible, if one is willing to assume that schools' actions are largely irrelevant to the patterns of who attends and graduates.

Graduation as a reflection of system dynamics

Green (1980) and Seidman (1996) have claimed, in contrast to focusing on the historical changes in labor markets, that school systems have powerful incentives to encourage greater educational attainment, up to a limit. The dynamics of educational systems, they claimed, push inevitably for more attainment. Arguing from the characteristics of an abstract labor-market, Green claimed that the marginal value of diplomas would be greatest when neither too low nor too high a proportion of students earned it. Thus, he claimed, incentive for attainment would grow dramatically between the low and high proportions, when the presumed value is highest. As each level of attainment becomes relatively saturated, the educational system will differentiate existing credentials by reputation, focus more on achievement than on attainment, and create a new norm for the next level of attainment. The theoretical ratcheting up of attainment standards is similar to Freeman's (1976) argument about overeducation and credentialism. The strength of Green's theory, and Seidman's application to graduation specifically, is its ability to explain the long plateau of high school graduation at the end of the twentieth century in a way that coincided with growing expectations for college and the appearance of a "dropout problem" (Dorn, 1993, 1996). As a much higher proportion of the population now expect to attend college for some time in contrast with 1960, the relative value of a high school diploma for most teenagers is as a stepping stone to college rather than an end in itself. In the meantime, Green asserted, high schools would be pressured to show more achievement for graduates (and less attention to merely graduating them), as the system shifts focus from attainment to differentiating by quality. .

This theory of internal systemic pressures has one primary weakness, the assumption that the educational system is the primary driver in educational attainment, generally without regard to changing labor-market conditions and changing views about the social mission of schools. Green describes his hypothesis about the benefits of an educational credential as a "tautological law" (Green, 1980, p. 94). The theory cannot explain the decline in graduation in the early 1940s, a matter which a reciprocal relationship between school and work explains with ease. In addition, the empirical claim of Green that the apparent advantages of attaining a diploma would be greatest when a moderate proportion of the population has attained it is inconsistent with evidence about relatively early advantages to high school attainment and a surprisingly high relative advantage in recent years (Boylan, 1993; Goldin, 1998). (Note 2) Green is, however, the most articulate proponent of the systematic explanation of broad trends in educational attainment, and his and Seidman's work reminds us that school systems have their own dynamics which will not always respond as desired to public policy.

Graduation as a reflection of high school custodialism

Others have suggested that a growing custodial mission for high schools is partly responsible for making graduation easier. Angus, Mirel, and Vinovskis (1988) pointed to education writings from the early twentieth century (Ayres, 1909; Thorndike, 1907) as prompting or marking a shift in public-school policies that encouraged schools' making work easier, de-emphasizing grade retention and academic coursework. Angus and Mirel (1999) argued, more comprehensively, that school officials deliberately crafted a custodial mission for secondary schools, in part out of lower expectations for the poorer students who were flooding high schools early in the century. They argued that the only way to change high schools' historical underestimation of adolescent ability to conduct academic work would be consequential tests tied to content standards. The custodial argument can explain both the general trends and also the dip in graduation during World War II. During most of the early twentieth century, this argument runs, the growing custodial mission of the schools cultivated a "shopping mall" institutional culture that made attendance and earning credits easier while discouraging hard work for the majority of students (Powell, Farrar, & Cohen, 1985). For a short time during World War II, keeping teens in school was less important than the war effort, but that change was temporary. Regular graduation statistics have remained on a plateau for the last few decades because the custodial mission concentrated on attendance, and graduation was an ancillary, if important, result for schools.

The custodial argument presumes a monolithic impulse to acquire and maintain enrollment for the legitimacy of the high school as an ideal of an institution. The custodial argument is weak where that impulse is clearly not monolithic. It cannot, by itself, explain the growth of alternative credentials over the past three decades. If the primary mission of high schools has successfully shifted from academics to custodialism, then there should be no path to a diploma reserved entirely for those *outside* regular attendance. Nor can the custodial argument explain why schools have regularly pushed out students (e.g., Fine, 1986, 1991), whom they should theoretically have an incentive to claim they successfully educate.

Crucial questions

In each theory described above, complex dynamics have been at work in determining high school attendance and graduation. The questions one may ask about the potential influence of high-stakes tests on graduation revolve around the relationships among various forces:

- Do high school students facing high-stakes tests calculate opportunity costs of further education differently from students not facing high-stakes tests?
- To what extent might the decisions of students to continue school change depending on existing labor-force conditions?
- What incentives do school systems have, stemming from within as well as without, that will affect how they respond to students who fail high-stakes tests?
- Is there independent evidence apart from test results that high-stakes testing has altered the low academic expectations high schools have often set in the past century, as argued by Angus and Mirel (1999), Powell, Farrar, and Cohen (1985), and Sizer (1984)?

As argued earlier, these explanations are not necessarily mutually exclusive. While the labor-market explanations suggest a largely external force, the other two theories suggest largely internal dynamics or decisions that shape the opportunities available to teenagers. We need not choose among these explanations at the moment, for they do not explicitly conflict. (Whether fellow researchers may prefer one explanation over another should not preclude consideration of them for policy purposes.) Each provides useful questions to frame further exploration. In particular, the theory of reciprocal labor market-school movement suggests that the effect of high-stakes testing may be indirect. Even if graduation gateway tests do not directly prevent diplomas for many, promotional gates earlier in school may result in a higher proportion of 18-year-olds who are far away from graduation. Will they stay in school at the expense of current earnings, if they will need to stay in school until 20? Advocates of these promotional and graduation gates point out that promoting and graduating students without skills are hollow events, and I have had enough students in my college classes without useful skills to be sympathetic with that argument. The assumption, discussed below, is that the value of a diploma is equivalent to the value of the knowledge and skills one presumably learns in school. However, a teenager is usually not learning academic skills if she or he leaves school for work. Historically, self-education has been a difficult, if virtuous, activity (Kett, 1994).

Stratified Diplomas

In addition to examining the history of changing attendance and graduation, one must also note, in a history of graduation, how schools have invented new ways to make distinctions among those who attend an institution. A diploma, originally, was a way to establish a category for a relatively small portion of the small group who attended high schools in the nineteenth century: those who completed a program of studies. The early twentieth century witnessed the introduction of additional diffferentiation through tracking and separate, specialized high schools. More recently, states have created different types of diploma. All of the changes in the last century form part of a regular institutional repertoire of ways of compartmentalizing students when pressed to solve specific problems. One may anticipate, based on this history, that all states may shortly invent new categories of diplomas for students who fail to meet exit exam requirements, if those numbers balloon, or that such students will be "cooled out" (see Clark, 1960) through non-academic exit routes.

For most of U.S. history, the meaning of a diploma was largely irrelevant to the value of education. The majority of those who attended higher education of any sort—colleges, academies, normal schools, grammar schools, and high schools—stayed briefly. Students and their families often found something of value in education even without the piece of paper documenting completion of a program of studies. For example, the vast majority of nineteenth-century students in Philadelphia's Central High School failed to earn a diploma. Despite this fact, high school attendance had become sufficiently attractive by the late nineteenth century to force Philadelphia's public schools to open more high schools (Labaree, 1988). So, too, those concerned about the inadequate education of children were more concerned with increasing experience (or what we might today call educational attainment) than with the acquisition of the diploma itself (Kett, 1995). Only in the 1960s (as described below) did the diploma become a sufficiently powerful expectation that its opposite, dropping out, commanded headline status as a social problem.

Even before the 1960s, though, the different potential goals for high school were creating an incentive to separate different groups of students in various ways. As Labaree (1988) explained with regard to Philadelphia, high schools were under pressure both to maintain the credential value of a high school education and also to open up access. The solution, reached in 1939 in Philadelphia, was a stratified set of high schools. Central High School was reborn as a selective high school open to students citywide through competitive admissions, with the other high schools as comprehensive high schools open to all in their attendance area. Many other large cities, such as New York, established similar, hierarchical organizations of schools. Some of the systems that did not have a Central High School or Bronx School of Science created vocational schools, which could serve to boost (or erode) the reputation of the other schools.

Differentiation through the establishment of high schools with unique programs and entrance criteria was one step beyond differentiation of students within high schools through tracking, which had developed earlier in the century (Angus & Mirel, 1999; Herbst, 1996; Krug, 1964). It was not a significant change in terms of how schools created separate expectations for different groups of students, though it was occasionally controversial locally, as in an attempt to create a separate vocational high school in Chicago early in the century (e.g., Katznelson & Weir, 1985). The creation of separate schools demonstrates, however, the way that public schools have been willing to create new programs in a flexible manner to respond to various pressures (Tyack & Cuban, 1995).

One consequence of this flexibility is that school systems have been willing to create programs for different levels of diplomas. In the last fifteen years, states across the country have created different official diplomas as well as an unofficial policy in some districts to steer students into GED programs. One recent report (Guy, Shin, Lee, & Thurlow,

1999) documented fifteen separate types of diploma in the fifty states and the District of Columbia (even after collapsing most diplomas into four basic categories). Three years ago, eight states had one type of exit document for high school students. Twelve states had some type of honors diploma, 35 had either "IEP" diplomas or certificates of attendance (typically available only for students with disabilities), and thirteen states had additional, idiosyncratic types of exit document.

Administrative databases (such as the Common Core of Data) often do not make such fine distinctions among types of diploma when granted by a single institution (public high schools, in this case). The proliferation of diploma types, both granted by public schools directly and also as alternatives to high school diplomas, requires some study. There is a real danger that some students, unable to succeed in mandatory academic diploma exams, would be granted alternative credentials that others would see as "watered-down," either directly through school programs or by steering into GED programs. The history of high schools has been replete with sometimes ingenious ways that high schools have undermined efforts to raise academic expectations for what students can do, and while one would wish that schools would avoid differentiation, one must not be blind to their tendency to engage in precisely this type of response to public policy demands.

Some might suggest that high school diplomas have become less important in their role for more people as an intermediate step to higher education than as a terminal degree. However, the changes in higher education, while parallel to high school differentiation, actually reinforce the importance of the high school diploma in itself, as the mid twentieth-century sequencing of high school and college schooling has eroded. One would be viewing the history of education too narrowly if one did not acknowledge that the differentiation of the high school (in terms of curriculum, buildings, and diplomas) has paralleled the differentiation of higher education. Colleges and universities developed electives and majors a century ago. More recently, discussion of community-college and adult vocational programs has focused on the alleged need for more credentials, certificates, and degrees (Parnell, 1985; National Center on Education and the Economy, 1990). Even though the broad push for vocational certificate programs envisioned by early Clinton administration officials has largely disappeared, community colleges have taken up the call for certificate-like programs on their own. At the same time as community colleges and high schools are creating new types of exit document, the century-long sequencing of schooling has started to decay. High school students can enroll in both high school and college in many places. A high proportion of college students are older than 25. Many adolescents and adults exit from and reenter schooling several times. Diplomas, thus, are no longer primarily an *exit* document for students who are leaving formal schooling. Understanding the impact of high-stakes testing on graduation requires consideration, therefore, of both the immediate value of a diploma for graduates and also its use as a key to formal schooling later in life. The significance of high school graduation has not degraded to the point of eighth grade graduation (for some proportion still have a high school diploma as the highest educational credential they will receive), but some parallel remains: both are necessary to continued attendance at school.

There are two logical consequences that follow from this broader context. First, measuring educational attainment requires a life-course approach to measuring schooling, perhaps not as complex as a multi-state life table (e.g., Land & Hough, 1989), but one that accommodates the various stops and starts in formal schooling that were common in the nineteenth century and are becoming more common today. Second, high school completion is a step in educational attainment with multiple uses. Measuring only the impact of high-stakes testing in terms of income misses the way that graduation credentials are prerequisites for higher education. If a consequence of high-stakes testing is a lower probability of graduating from high school, then a long-term result might be a lower probability of having access to quality higher education later in life.

Meaning

The recent debate over how standardized testing may affect graduation chances for students demonstrates how the social meaning of a diploma has changed over the past century. In 1900, high school graduation was rare, and the act of leaving school before attending high school or earning a diploma was an expected, if sometimes lamented, fact of life for most teenagers. If standardized testing had existed for graduating high school students then, probably very few would have been concerned about the potential effects on students. One hundred years later, those who leave school without a diploma are violating a normative expectation about what adolescents do (attend and graduate from high school), and many deem that violation sufficiently dangerous to society as a whole that the action has a special term—"dropping out"—and often earns public-policy scrutiny. In part, we worry about dropouts for economic reasons, though there is considerable debate about the extent to which a diploma represents additional human capital rather than the competitive value of a credential in a labor queue (e.g., Becker, 1964; Berg, 1970; Dore, 1976).

Those in the U.S. have worried about dropping out as a social problem, for the popular image of a high school dropout is of an adult without a viable future, an emblem of dependence (Dorn, 1996). Throughout the last century, educators and others have occasionally argued for public-policy remedies to the act of leaving school, and beginning in the 1960s, the "dropout problem" became the focal point of deliberate, well-publicized institutional efforts. Those debates have affected the larger patterns of attendance and graduation very little, however. The fact that specialized programs other than the GED have not seemed to affect the larger trends suggests that public-policy discussion of dropping out has been ineffective. The greatest expansion in the rights of students to attend or continue attending schools came in the 1970s, long after the height of the first wave of headlines over dropping out and generally unconnected to it. Too often, public policy discussion of dropping out is disconnected from the larger patterns of school system behavior. One should thus be wary of narrow interpretations of the graduation-test question that may omit crucial features of schools. In particular, the proliferation of diploma types, in addition to the act of leaving school, is a crucial new feature of secondary education in the U.S. that deserves attention.

The Recent Norm of Graduation

Debate about dropping out in the past four decades reflects the relatively new expectation of graduation for teenagers. Impossible before most adolescents attended high school for several years, this expectation bloomed during the 1960s into a public, headline-grabbing discussion about why we should be concerned about dropping out and what schools might do to solve the problem. Two aspects about the development of graduation as a norm (or expectation) are important to the discussion at this workshop. First, graduation is an *age-related* norm, part of the historical growth in age consciousness that has paralleled the development of retirement and other "stages of life" that did not exist in public awareness two centuries ago (Chudacoff, 1989; Graebner, 1980; Haber, 1983; Kett, 1977). Because traditional high school programs serve teenagers, and because most adult education programs for dropouts focus on the GED as a goal, *in most cases only teenagers can meet that norm of graduation. Also, graduation statistics that focus on the teen years* most appropriately measure the extent to which high schools meet that norm.

In addition to expecting graduation as a part of adolescence, many in the U.S. also expect graduation to solve many potential problems facing both individual teenagers and society in general. Many argue for attention to dropping out because high school graduates earn more, are less likely to be in jail, are less likely to have children out of wedlock, and so forth. As sociologist Lucius Cervantes asserted more than three decades ago, today's dropouts would be the "gangsters, hoodlums, drug addicted, government-dependent-prone, irresponsible and illegitimate parents of tomorrow" (Cervantes, 1965, p. 197). The dropout literature in the 1960s married the tangible economic penalties of dropping out (higher unemployment and lower income) with the assumption that dropouts were psychologically weak, delinquency prone males to create a stereotype of dropouts as those who would, in the future, be dependent on society (Dorn, 1993, 1996).

The 1960s were certainly not the first time that either educators or school critics were concerned about how students attended or left school, nor was it the first decade when anyone claimed that schools should address a multitude of social problems. Educators and social critics have claimed, at various times since the establishment of English colonies in North America that schools should train future leaders, assimilate immigrants, discourage immorality among the poor, prevent class conflict, inculcate nationalistic values, and improve the character of workers, to name some of the goals. However, for most of U.S. history, sporadic attendance—not leaving school before acquiring a diploma—has been the target of reformers' efforts (Kaestle, 1983; Tropea, 1987). Even when some have been concerned about how much educational experience a child has received (as opposed to its consistency), the goal has been general educational attainment rather than the specific target of high school graduation (Kett, 1995).

What was new in the 1960s was the link between older beliefs about the value of schools, on the one hand, and the new expectation of graduation, on the other. Symbolizing this link was a word that, before the 1960s, had been one of several ways that educators talked about those who left school before high school graduation. Student withdrawals and early school leavers were certainly the topics of debate from 1900 through the late 1950s. In the past four decades, however, those who discussed student attrition have generally used one word, *dropout* (or variants of it), to name the problem. Dropping out has become the inverse of graduation, representing individual and social danger and growing directly from our new expectation that teenagers graduate from high school. The second national educational goal, 90 percent high school graduation (however measured), culminates a sea change in how we as a society have been concerned about attendance. The creation of an expectation for graduation was predictable once the majority of teens began graduating (also see Green, 1980). However, the form that the expectation took and the cultural rationale for it were not. As a society, we have chosen to be concerned about dropping out as a cause of adult dependence. Thus, many warnings about the long-term consequences of barriers to graduation today take the form of social results like crime and unemployment. Raising a discussion focused on alternative concerns (such as equity) is difficult. And, in general, disentangling all the issues is harder because we have a legacy of public debate that has conflated them. (Labaree, (1997)

Ineffective Public Policy

Deliberate public policy efforts to reduce dropping out have suffered from limits in scope, contradictions in the purposes of high schools, the prevailing stereotypes about dropouts, and a persistent belief in the power of public relations. One needs to be cautious about generalizations because dropout policies vary by local district and also by government level (district, state, and federal). Local circumstances have shaped most dropout programs, even where funds flow from above. However, one must contrast the pronouncements regularly made about dropout prevention with the relative stability in regular high school graduation over the past few decades (the time when we have described school attrition as *dropping out*). While an individual program may well have helped teenagers in it (and many programs certainly have), dropout prevention programs have, together, not changed the nationwide patterns of graduating and leaving school. The obvious question is, why have programs had such little effect overall?

The obvious answer—but an important one—is that dropout prevention programs have generally been small, focusing on a few students or dropouts at a time. They are *programs* rather than *changes in schools*. Tyack and Cuban (1995) argue that schools have, historically, been far more willing to adopt small, incremental changes than large ones. A school system can more easily approve one, two, or ten small dropout prevention programs than it can change prevailing expectations teachers and principals may have about students. New York City's public schools, which had the most programs explicitly labeled "dropout prevention" in the 1960s, never planned (even at the beginning of the War on Poverty) to serve more than a fraction of adolescent dropouts (Dorn, 1996, p. 88). Dropout prevention efforts are, while sometimes different in character from the earliest ones, still not very different in scope. They still involve counseling, still have difficulties with funding, and still are on the edges of school systems' organizations (Dynarski & Gleason, 1998). School systems like New York City often eliminated dropout prevention efforts when outside funding dried up in the later 1960s, and one can still hear similar tales today. Systematic dropout prevention is not a high priority for most local school systems.

One reason why dropout prevention is not a priority is that it contradicts an abiding incentive for public school systems to restrict credentials. Even if one disagrees with Green's (1980) and Seidman's (1996) argument about the limits of educational credentials, there is a long history of high schools being rewarded for either restricting all high school credentials or stratifying them and restricting the most valued ones (Labaree, 1988). As discussed in the **Diplomas** section above, public school systems have created selective-admission high schools and programs, and these features of the system often generate the most positive news for public schools. Public high schools thus have contradictory missions, attempting both to educate all adolescents and also to provide the highest rewards to a limited few. As part of that structure and the rhetoric of meritocracy in North American society in general (Lemann, 1999), many teachers and principals believe deeply in schooling as a meritocratic system. The administrator who told Fine (1986, 1991) that low-performing or troubling students were "hijackers" of his school was not alone. Many educators simply do not want to educate everyone. Even the most well-meaning teachers often resort to a form of educational triage, trying to save a few children while tolerating the failure of others (e.g., Michie, 1999; Sapon-Shevin, 1993; Sizer, 1984).

In addition to being small and fighting competing goals for public schools, the broad assumption of psychological problems of potential and real dropouts has encouraged programs that focus on remedying individual defects rather than addressing needs more broadly. Consistent with the common belief that dropouts were budding juvenile delinquents on the way to unemployment with little thought of their futures, many early dropout programs featured either individual counseling (to solve personal adjustment issues) or work experience programs (to increase the chances of employment later) (Dorn, 1996). I know of no dropout program in the 1960s that addressed concrete needs in a narrow fashion, such as providing day care services for students with children. Even now, providing day care is an exception rather than the norm as part of dropout prevention strategies (see Orr, 1987, for a smorgasbord approach to dropout prevention). Focusing on student deficits is certainly not unique either to dropout prevention or to the United States; critics of special education or other at-risk programs often point out deficit orientation, and the former Soviet Union had an Institute on Defectology (see Vygotsky, 1993, for the casual use of the word, akin to the English "handicapped"). The deficit focus of dropout prevention has had two consequences, one practical and one political. The practical consequence of a deficit orientation is that such programs may not solve the problems dropouts have that are outside their control (such as poverty, inadequate public transportation, or school-caused problems). The political consequence of a deficit orientation is that such programs have few lasting constituencies that can fight for their long-term survival , as Cuban (1992) describes is necessary for the longevity of school reform efforts (Dorn, 1996).

Lastly, nationally visible dropout prevention efforts have consistently assumed the persuasive powers of public relations, following a pattern that dates back to before dropping out became a headline issue. The two World Wars, the 1950s, 1963, and the early 1990s all witnessed national public-relations efforts (generally through public-private partnerships) to convince dropouts to return to school or students to remain in school. Four of the five efforts went by the same name: the "Stay in School" campaign (Angus, 1965; Dorn, 1996). None had a documented meaningful effect, because public relations efforts could not significantly change the reasons why most dropouts would leave school. In particular, the push and pull influences described above have generally gone unmentioned in public policy debate. The assumption that public relations alone can change broad demographic patterns is simplistic.

Limits on Public Policy Debate

Consistent with the limits of explicit dropout prevention and remediation efforts have been limits on public policy debate. We do not always talk about practices that are responsible for more attrition than small programs can ever compensate for. The first wave of explicit dropout programs in the 1960s demonstrated this type of omission. Despite arguments about the importance of eliminating dropping out, neither schools nor their critics suggested changing three crucial policies that encouraged dropping out: the exclusion of students with disabilities from schools, the regular separation of pregnant teenagers from schools, and widespread suspension and expulsion without due process. Challenges to all of those practices succeeded in the 1970s, well after the initial peak of concern over dropping out and by using a civil-rights argument rather than highlighting the social costs of under-education (Dorn, 1996). While discussion in the 1980s was more likely to focus on the civil-rights dimensions of student attrition, public policy discussion still has revolved around the perceived link between dropping out and dependency. Even liberals like California Rep. George Miller have justified federal aid in a way that 1960s writers on the dropout problem would have recognized:

> Without a high school education, few will be able to compete in the new, high-technology centered labor market. Dropout prevention programs are essential to securing family self-sufficiency and to prevent the cycle from starting over again with a new generation of children. (Miller, 1987, p. H3901-2)

Discussion of dropping out as an equity concern has certainly existed in the last fifteen years, but it has not dominated public debate (Dorn, 1996). The history suggests that the shadow of dependency will tend to dominate discussion of dropping out. Those trying to persuade the public to pay attention will emphasize the social costs of dropping out, echoing the last 35 years of writings and speeches on dropping out. In trying to get attention, these appeals to fears about dependency can easily drown other ways of framing dropping out.

Shifting Tensions

The tensions described here, between broadening educational access and restricting credentials, have shifted in the past forty years from its roots in schools as institutions to broader political debate. In 1960, many of those who had doubts about the wisdom of graduating everyone would have been older educators born before 1920, who had gone to school when only a minority graduated from high school. The changes in law in the 1970s limiting the ability of schools to exclude students, as well as the stable proportion (a large majority) who do graduate from high school, have been in place now for more than a quarter of a century. Most educators today grew up when graduation was a norm and have been professionals, for the most part, during a time when schools have had to open their doors to all children. Today, the institutional tension in the mission of high schools remains. However, I see far more tension in public debate than in schools as organizations. Policymakers simultaneously call for 90 percent graduation (in the second national goal) and an end to social promotion. Advocates of high-stakes testing argue that testing with consequences will provide motivation for teachers and students to work hard and thus accomplish universal high-quality education. The push for high-stakes testing is, in part, a consequence of that shift to broader political discussion about schooling in a national debate. Many states have taken this road to what seems, to this historian, a holy grail of testing (Goldstein, 1997). The motivation of many is admirable; however, introspection about the dilemmas of schooling is largely absent.

Lessons and Perspectives

The history of graduation, diplomas, and concerns about dropping out in the last century provides a guide, if not to the future, at least to some plausible issues of concern in the discussion of graduation in an era of high-stakes testing. These concerns mirror those of the National Research Council's (2001) Committee on Educational Excellence and Testing Equity. There is a real possibility that increased use of high-stakes testing will decrease in some measurable amount the likelihood of teens graduating from regular high school programs in the U.S. in general. The recent stability in the regular high school graduation rate is evidence that graduation does not always increase, and the decline at the end of the World War II is evidence that it can decrease under certain circumstances. One could imagine, plausibly, that the direct and indirect effects of high-stakes tests might result in a lower likelihood of teens' graduating with academic diplomas. In addition, the recent proliferation of high school credentials, together with the history of high school differentiation, should raise the empirical question of "*which* diploma earned" to a high priority in any research tracking the consequences of high-stakes testing. Finally, the small size of most dropout programs historically and the blind spots in public discussion of dropping out should warn anyone against the belief that a few band-aid programs might easily staunch a large flow of teenagers heading through school doors. These issues raised by the history described here do not predict whether high-stakes tests will restore some presumed value of the diploma or increase the likelihood of student attrition. They should, however, help frame further discussion and research.

Moreover, the long tangled history of diplomas and the different possible values of a high school degree suggests that we must be very clear when discussing the rationale for high-stakes tests. Diplomas have an exchange value, either in a labor market or to gain entry to another school, and that exchange value may not be directly related to the purpose of a degree requirement. On what basis are states withholding a standard academic diploma from students who do not meet certain examination requirements? Having the diploma for its exchange value is important to most students, and so one must balance the property interest in a diploma (as the *Debra P.* case described it) against the eventual purpose of the degree requirement. When is threatening a student's access to the exchange value of a diploma a justifiable policy? Consider the following potential goals of any requirement:

- Improving the education of the student directly;
- Improving the exchange value of the diploma for the student (perhaps by increasing its credibility immediately with employers or colleges);
- Certifying concrete skills or knowledge in an educational program for society (where a degree requirement is tied to the specific program);
- Certifying the general intellectual worth of graduates for society (where a degree requirement is not tied to a specific program);
- Improving the status of a school or schools (through the credibility of a degree's inherent worth or through restricting credentials);
- Improving the exchange value of the diploma for future students (by increasing its credibility in the long term); and
- Using the pressure from current failures to encourage schools to improve in the future.

I have ordered these seven goals from the most immediate to the most distant in relationship to the students. In the first two goals, there are alleged direct benefits to the student, either in the use value of a diploma (the learning) or in the exchange value of the diploma. In the third and fourth goals, with social benefits, the student may gain indirectly. The last three goals use the student as a means, in essence, to others' ends, with little foreseeable benefit to the students in question. One may evaluate some of these goals in terms of ethics. I would, for example, disapprove of policies where the goal uses current students primarily as pawns. Other goals, such as the first two, are amenable to empirical exploration. The third goal is most common in professional preparation programs, where the state has a clear interest in skilled graduates of medical schools, colleges of education, and the like, and where one might directly evaluate the relevance of specific requirements.

The hardest goal to weigh against the property interest in a diploma is certifying graduates as educated *in toto*. General education requirements, whether they are courses or exams, are always proxies for evaluating whether a student is sufficiently skilled or intellectual to have earned the title "graduate." The risk in either type of requirement is that passing a course or exam might be, at best, a poor substitute for being a well-educated graduate. If one can narrow the goal of graduation gateway testing to this general certification, though, one can examine the merits of the proxy relationship more clearly. Would the expectations be reasonable (or a healthy intellectual stretch) for a well-educated adult today? Do the exams measure what adults value in a well-educated neighbor? Do all students have a clear opportunity to learn the skills and knowledge by the end of a phase-in period?

We rarely discuss this proxy relationship in a straightforward manner because of the social and political baggage that the high school diploma has acquired over more than a century. We assume, all too often, that the skills and knowledge learned before, the exchange value of, and the social benefits of diplomas are identical. As high school graduation has become common, and as dropping out has acquired the connotation of dependency, we have focused many of our anxieties about a changing world on teenagers and a piece of paper that most—but not all—acquire. This historian believes firmly that disentangling all these social meanings of diplomas would help us analyze the potential consequences of high-stakes tests for teenagers, the meaning of those consequences, and how to explain this analysis to a country that is desperate for a diploma with some recognized meaning.

Notes

1. The series of the diploma-to-17-year-old-population ratio mixes data from Goldin (1999) with more recent and updated information from the U.S. Department of Education (Snyder & Hoffman, 2001). Since teenagers are now more likely to graduate at age 18 than in most of the twentieth century, that ratio based on the 17-year-old population will underestimate graduation when the 17-year-old population is larger than the 18-year-old population and will overestimate graduation when the 18-year-old population is larger. In neither will the differences be meaningful in terms of the long-term trends described here, though the measure is not a true event-exposure rate or a probability (in the same way that the infant mortality rate is not a true event-exposure rate but nonetheless is a rough measure of mortality conditions at the beginning of life). The interval spans for both the Dorn (1996) and Kaufman et al. (2001) series are mapped back to the birth cohort of the mid-point of the range: The 18-24-year-old graduation rate reported for 2000 by Kaufman et al. is mapped back to the birth year of those age 21, or 1979. It is, in effect, a step to smooth the estimate. As Goldin (1998) noted, adults exaggerate when self-reporting education. The census in 1940 experienced far more evident exaggeration (when compared to graduation statistics Goldin gathered) than more recently, so the growing spread between the self-reported figures and the diploma-to-17-year-old population ratio in the past few decades is not likely to be a result of *increased* exaggeration by those surveyed.

2. Green (1980) and Seidman (1996) have argued that the standard assumption of a relative advantage of educational attainment (comparing having a degree to not having it) ignores the way that creeping attainment in a cohort make the relevant comparison the mean proportion of the cohort. As a higher proportion of a cohort have a degree, the relative advantage of having a diploma (compared to the rest of one's cohort) disappears, and the more visible issue becomes the penalty of missing that diploma (relative to the cohort as a whole). Whether such a distinction is meaningful in economic terms is an open question, but the socially-perceived advantages and penalties, however, are often in comparison with relative norms. In that regard, Green and Seidman are substantially correct: high school graduation has become more commonly a stepping-stone to college since 1965, especially for children of wealthier families, while dropping out is now universally perceived as a serious impediment to economic success for adults.

References

Angus, D. (1965). The dropout problem. Unpublished Ph.D. diss., Ohio State University, Columbus, OH.

Angus, D., & Mirel, J. E. (1999). *The failed promise of the American high school, 1890-1995*. New York: Teachers College Press.

Angus, D., Mirel, J. E., & and Vinovskis, M. A. (1988). Historical development of age stratification in schooling. *Teachers College Record*, 90, 211-236.

Ayres, L. P. (1909). *Laggards in our schools*. New York: Survey Associates, Inc.

Becker, G. (1964). *Human capital*. New York: Columbia University Press.

Berg, I. (1970). *Education and jobs: The great training robbery*. New York: Praeger.

Boesel, D., Alsalam, N., & Smith, T. (1996). *Educational and labor market performance of GED recipients*. Washington, DC: National Library of Education.

Boylan, R. D. (1993). The effect of the number of diplomas on their value. *Sociology of Education*, 66, 206-221.

Cameron, S. V., & Heckman, J. L. (1993). The nonequivalence of high school equivalents. *Journal of Labor Economics*, 1, 1-47.

Cervantes, L. F. (1965). *The dropout*: Ann Arbor, MI: University of Michigan Press.

Chandler, A. D. (1977). *The visible hand*. Cambridege, MA: Belknap Press.

Chudacoff, H. P. (1989). *How old are you?* Princeton, NJ: Princeton University Press.

Clark, B. R. (1960). The "cooling out" function in higher education. *American Journal of Sociology*, 65, 569-76.

Clinton, W. J. (1999). State of the Union address. Retrieved October 2, 2001, from http://clinton3.nara.gov/WH/S OTU99/

Cremin, L. A. (1989). *Popular education and its discontents*. New York: Harper & Row.

Cohen, M. (1992). *Workshop to office*. Ithaca, NY: Cornell University Press.

Cuban, L. (1992). Why some reforms last. *American Journal of Education*, 100, 166-194.

Davies, M. (1982). *Woman's place is at the typewriter*. Philadelphia: Temple University Press.

Debra P. v. Turlington, 644 F. 2d 397 (5th Cir. 1981), 564 F. Supp. 177 (M.D. Fla. 1983).

DeMott, B. (1990). *The imperial middle: Why American's can't think straight about class*. New York: Morrow.

DeSanctis, V. (1979). *The Adult Education Act 1964-1979*. Upper Montclair, NJ: National Adult Education Clearinghouse. (ERIC Reproduction Document No. ED 218 517).

Dore, R. P. (1976). *The diploma disease: Education, qualification, and development*. Berkeley: University of California Press. Dorn, S. (1996). *Creating the dropout: An institutional and social history of school failure*. Westport, CT: Praeger.

Dorn, S. 1993. Origins of the "dropout problem." *History of Education Quarterly*, 33, 353-373.

Dorn, S. 1996. *Creating the dropout: An institutional and social history of school failure*. Westport, CT: Praeger.

Dorn, S., & Johanningmeier, E. V. 1999. Dropping out and the military metaphor for schooling. *History of Education Quarterly, 39*, 193-198.

Dynarski, M., & Gleason, P. (1998). *How can we help? What we have learned from evaluations of federal dropout-prevention programs*. School Dropout Demonstration Assistance Program Evaluation Research Report. Princeton, New Jersey: Mathematica Policy Research, Inc. Retrieved June 23, 2000, from http://www.dropoutprevention.org/2levelpages/downloads/dod- syn.pdf

Ekstrom, R. B., Goertz, M. E., Pollack, J. M. , & Rock, D. A. (1986). Who drops out of high school and why? *Teachers College Press*, 87, 357-373.

Featherman, D. L., & Hauser, R. M. (1976). Equality of schooling. *Sociology of Education*, 49, 99-120.

Fine, M. (1986). Why urban adolescents drop into and out of high school. *Teachers College Press*, 87, 393-408.

Fine, M. (1991). *Framing dropouts*. Albany, NY: State University of New York Press.

Freeman, R. (1976). *The overeducated American*. New York: Academic Press.

GI Forum v. Texas Education Agency, 87 F. Supp. 2d 667 (2000).

Glass, G. V & Edholm, C. A. (2002). The AIMS test and the mathematics actually used by Arizona employees. EPSL-0209-119 EPRU. Tempe, AZ: Education Policy Studies Laboratory, Arizona State University, available on-line at http://www.asu.edu/educ/epsl/EPRU/documents/EPSL-0210-122-EPRU.html.

Goldin, C. (1999). A brief history of education in the United States. NBER Historical Paper H0119. Cambridge, MA: National Bureau of Economic Research, available on-line at http://papers.nber.org/

Goldin, C. (1998). America's graduation from high school: The evolution and spread of secondary schooling in the twentieth century. *Journal of Economic History*, 58, 345-374.

Goldstein, H. (1997). Value added tables: The less-than-holy grail. *Managing Schools Today*, 6(6), 18-19.

Graebner, W. (1980). *A history of retirement*. New Haven, CT: Yale University Press.

Green, T. F. (1980). *Predicting the behavior of the educational system*. Syracuse, NY: Syracuse University Press.

Greenberger, E., & Steinberg, L. (1986). *When teenagers work*. New York: Basic Books.

Grissmer, D. W., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us*. Santa Monica, CA: RAND, report MR-924- EDU.

Guy, B., Shin, H., Lee, S.-Y., and Thurlow, M. L. (1999). *State graduation requirements for students with and without disabilities*. Technical Report 24. Minneapolis, MN: National Center on Educational Outcomes. Retrieved June 22, 2000, from http://education.umn.edu/NCEO/OnlinePubs/Technical24.html

Haber, C. (1983). *Beyond sixty-five*. New York: Cambridge University Press.

Hamilton, S. F. (1987). Raising standards and reducing dropout rates. In G. Natriello (Ed.), *School dropouts*. New York: Teachers College Press.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). Retrieved October 2, 2001, from http://epaa.asu.edu/epaa/v8n41/

Herbst, J. (1996). *The once and future school*. New York: Routledge.

Jaynes, G. D, & Williams, R. M., Jr., (Eds). (1989). *A common destiny*. Washington, DC: National Academy Press.

Kaestle, C. F. (1983). *Pillars of the Republic*. New York: Hill and Wang.

Kaestle, C. F., & Vinovskis, M. A. (1980). *Education and social change in nineteenth-century Massachusetts*. New York: Cambridge University Press.

Kantor, H. A. (1983). Vocationalism in American education. In H. A. Kantor & D. Tyack (Eds.), *Work, youth and schooling*. Stanford, CA: Stanford University Press.

Katz, M. B., Doucet, M. J., & Stern, M. J. (1982). *The social organization of early industrial capitalism*. Cambridge, MA: Harvard University Press.

Katznelson, I., and Weir, M. (1985). *Schooling for all*. New York: Basic Books.

Kaufman, P., Alt, M. N., & Chapman, C. (2001). *Dropout rates in the United States: 2000*. Washington, DC: National Center for Education Statistics.

Kaufman, P., Kwon, J. Y., Klein, S., & Chapman, C. D. (2000). *Dropout rates in the United States: 1999*. Washington, D.C.: National Center for Education Statistics.

Kett, J. K. (1977). *Rites of passage*. New York: Basic Books.

Kett, J. K. (1994). *The pursuit of knowledge under difficulties: From self-improvement to adult education in America, 1750-1990*. Stanford, CA: Stanford University Press.

Kett, J. K. (1995). School leaving. In D. Ravitch & M. A. Vinovskis (Eds.), *Learning from the past*. Baltimore: Johns Hopkins University Press.

Klein, S. P., Hamilton, L. S., McCaffrey, D. F, & Stecher, B. M. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives*, 9 (49). Retrieved October 2, 2001, from http://epaa.asu.edu/epaa/v8n49/.

Kominski, R. (1980). Estimating the national high school dropout rate. *Demography*, 27, 303-312.

Kossan, P. (2000, November 22). Keegan backs off AIMS requirements. *Arizona Republic*, p. 1.

Kreitzer, A. E., Madaus, G. F., & Haney, W. (1989). Competency testing and dropouts. In L. Weis, E. Farrar, & H. G. Petrie (Eds.), *Dropouts from school*. Albany, NY: State University of New York Press.

Krug, E. A. (1964). *The shaping of the American high school* (2 vols.). New York: Harper & Row.

Labaree, D. F. (1988). *The making of an American high school*. New Haven, CT: Yale University Press.

Labaree, D. F. (1997). *How to succeed in school without really learning: The credentials race in American education*. New Haven: Yale University Press.

Land, K. C., & Hough, G. C. Jr. (1989). On the combination of prevalence rate and increment-decrement methods for tables of school life, with applications to the 1969-70, 1974-75, and 1979-80 school years. *Journal of the American Statistical Association*, 84, 63-75.

Lemann, N. (1999). *The big test*. New York: Farrar, Straus, and Giroux.

Lieberson, S. (1980). *A piece of the pie*. Berkeley: University of California Press.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.

Madaus, G., (Ed.). (1983). *Thecourts, validity, and minimum competency testing*. Boston: Kluwer-Nijhoff Publishing.

Mare, R. D. (1980). Social background and school continuation decisions. *Journal of the American Statistical Association*, 75, 295-305.

Mare, R. D. (1981). Change and stability in educational stratification. *American Sociological Review*, 46:, 72-87.

McDill, E. L., Natriello, G., & Pallas, A. M. (1985). Raising standards and retaining students: The impact of the reform recommendations on potential dropouts. *Review of Educational Research*, 55, 415-433.

McDill, E. L., Natriello, G., & Pallas, A. M. (1987). A population at risk: Potential consequences of tougher school standards for student dropouts. *American Journal of Education*, 94, 135-181.

Michie, G. (1999). *Holler if you hear me*. New York: Teachers College Press.

Miller, G. (1987, May 21). *Congressional Record*.

Murnane, R. J. , Willet, J. B., & Boudett, K. P. (1995). "Do High School Dropouts Benefit from Obtaining a GED?" *Education and Policy Analysis*, 17, 133-147.

Murnane, R. J., Willet, J. B., & Tyler, J. H. (1999). *Who Benefits from Obtaining a GED? Evidence from High School and Beyond*, NBER Working Paper No. W7172. Cambridge, MA: National Bureau of Economic Research, available online at http://papers.nber.org/

National Center for Education Statistics. (2001). *Digest of educational statistics, 2000*. Washington, DC: Government Printing Office.

National Center on Education and the Economy. (1990). *America's choice: High skills or low wages!* Rochester, NY: National Center on Education and the Economy.

National Research Council. (2001). *Understanding dropouts: Statistics, strategies, and high-stakes testing*. Washington, DC: National Academy Press.

Orr, M. (1987). *Keeping students in school*. San Francisco: Jossey-Bass.

Ossowski, S. (1963). *Class structure in the social consciousness*. London: Routledge and Kegan Paul.

Osterman, P. (1980). *Getting started*. Cambridge, MA: MIT Press.

Parnell, D. (1985). *The neglected majority*. Washington, DC: Community College Press.

Perlmann, J. (1988). *Ethnic differences*. New York: Cambridge University Press.

Powell, A., Farrar, E., & Cohen, D. (1985). *The shopping mall high school*. Boston: Houghton Mifflin.

Pullin, D. (1981). Minimum competency testing and the demand for accountability. *Phi Delta Kappan*, 63(1), 20-22.

Roderick, M. R. (1993). *The path to dropping out*. Westport, CT: Auburn House.

Rumberger, R. W. (1987). High school dropouts. *Review of Educational Research*, 57, 101-121.

Rumberger, R. W. (1998). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal*, 32, 583-625.

Rumberger, R. W., & Larson, K. A. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education*, 107, 1-35.

Rumberger, R. W., & Scott, T. L. (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education*, 73, 39-67.

Sapon-Shevin, M. (1993). Gifted education and the protection of privilege: Breaking the silence, opening the discourse. In L. Weis and M. Fine (Eds.), *Beyond silenced voices: Class, race, and gender in United States Schools* (pp. 25-44). Albany, NY: State University of New York Press.

Salganik, L. H. (1985). Why testing reforms are so popular and how they are changing education. *Phi Delta Kappan*, 66, 607-610 .

Seidman, R. H. (1996). National education 'Goals 2000': Some disastrous unintended consequences. *Education Policy Analysis Archives*, 4(11). Retrieved October 2, 2001, from http://epaa.asu.edu/epaa/v4n11/

Sizer, T. (1984). *Horace's compromise*. Boston: Houghton Mifflin.

Snyder, T. D., (Ed.). (1993). *One hundred twenty years of American education*. Washington, DC: National Center for Education Statistics.

Snyder, T. D., & Hoffman, C. M., (Ed.). (2002). *Digest of education statistics 2001*. Washington, DC: National Center for Education Statistics.

Thorndike, E. L. (1907). *The elimination of pupils from school*. U.S. Bureau of Education Bulletin No. 4, Whole No. 379. Washington, DC: Government Printing Office.

Tropea, J. (1987). Bureaucratic order and special children: Urban schools, 1890s-1940s. *History of Education Quarterly*, 27, 29-53.

Turlington, R. D. (1985). How testing is changing education in Florida. *Educational Measurement: Issues and Practice*, 4(2), 9-11.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1993). *The collected works of L. S. Vygotsky, Vol. 2: The fundamentals of defectology (abnormal psychology and learning disabilities)* (J. E. Knox & C. B. Stevens, Trans.). Cambridge, MA: Perseus Publishing.

Walters, P. B., & Briggs, C. M. (1993). The family economy, child labor, and schooling. *American Sociological Review*, 58, 163-181.

Acknowledgment

This article began as a commissioned work for the July 2000 workshop on school completion in standards-based reform of the National Research Council's Committee on Educational Excellence and Testing Equity. I am grateful for comments by an anonymous reviewer for the *Education Policy Analysis Archives* as well as by Robert Hauser, Ulric Neisser, Diana Pullin, Bella Rosenberg, Russell Rumberger, and William Trent that shaped revisions, as well as the permission of the National Research Council to use that commissioned paper as the basis for this article. Any opinions expressed are those of the author and neither those of the persons mentioned above nor of the National Research Council.

About the Author

Sherman Dorn

University of South Florida EDU162
4202 E. Fowler Avenue
Tampa, FL 33620-7750
E-mail: sdorn@tempest.coedu.usf.edu

Sherman Dorn is Associate Professor in the Department of Psychological and Social Foundations in the College of Education at the University of South Florida. He teaches social foundations courses, and his interests include how schools have historically treated marginalized populations and the construction of policy problems in educational politics.

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires

javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@airbrasil.org.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

This article has been retrieved 4405 times since January 11, 2003

Education Policy Analysis Archives

Volume 11 Number 2 **January 11, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article if EPAA is credited and copies are not sold. EPAA is a project of the Education Policy Studies Laboratory.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Policymakers' Online Use of Academic Research

John Willinsky
University of British Columbia

Citation: Willinsky, J. (2003, January 11). Policymakers' online use of academic research, *Education Policy Analysis Archives, 11* (2). Retrieved [date] from http://epaa.asu.edu/epaa/v11n2/.

Abstract

In addressing the question of how new technologies can improve the public quality and presence of academic research, this article reports on the current online use of research by policymakers. Interviews with a sample of 25 Canadian policymakers at the federal level were conducted, looking at the specific role that online research has begun to play in their work, and what frustrations they face in using this research. The study found widespread use of online research, increasing the consultation of this source in policy analysis and formation. The principal issues remain those of access, indexing and credibility, with policymakers restricting themselves in large part to open access sources. Still, online research is proving a counterforce to policymakers' reliance on a small number of academic consultants as gatekeepers and sources for research. What is needed, it becomes clear, is investigations into whether innovative well-indexed systems that integrate a range of academic and non-academic resources might increase the political impact of research in the social sciences and education.

Introduction

To better understand the complex and diffuse relationship that exists between social policy and social science research, I present the reflections of twenty-nine Canadian policy officials on access to scholarly research afforded by the Internet. While researchers interested in the political impact of the Internet have tended to focus on "digital democracy" issues of public access to government information, public consultation and participation, and public privacy and surveillance, a far less dramatic change is also taking place to existing policy-making processes. (Note 1) The knowledge economy of the Internet has significantly increased policymakers' ability to tap into current, critical, and relevant research without leaving their desk. Although many academic journals remain closed to those without access to a good research library, a growing number of individual papers, journals, and research archives are providing full online access at no charge to users. (Note 2) This new availability has transformed the information environment within which policy development takes place, and it is altering the role of social scientists in democratic processes.

Admittedly, the influence of academic research on government policy pales before the impact of the government's own sources of information, whether from politicians, bureaucrats, or government professionals. As Harvard's Carol Weiss has observed, it takes "an extraordinary concatenation of circumstances" for "research to influence policy decisions directly" (1991, p. 44). Weiss has concluded, after many years of evaluation work in education, that "governments don't often use research directly, but research helps people reconsider issues, it helps them think differently, it helps them reconceptualize what the problem is and how prevalent it is, it helps them discard some old assumptions, it punctures old myths. It takes time and reconceptualization before research actually leads to a change in policy" (1998). Still, one finds, for example, that a recent U.S. federal education bill calls on school officials who seek federal support to secure "scientific evidence" in arriving at program decisions, as well as related initiatives to improve federal spending on education research. (Note 3) Similarly, the British government's recent Hillage report calls for education research to play a greater part in policy formation, and to take steps to better prepare itself to play this role (Hillage 1998). Also, the University of London's Institute for Education has established a Centre for Evidence-Informed Policy and Practice, which seeks, among other things, to make research more accessible for policymakers and the public, as well as to support a more coordinated, less fragmented, approach to education research. (Note 4)

Given the complex and exhaustive knowledge needs of the modern democratic state, academic research has a vital, irreplaceable role to play in not only informing and evaluating government decisions, programs and policies, but also in a broader intellectual sense, whether by challenging assumptions, providing a historical perspective, bringing international comparisons to bear, or offering lessons from experimental models and alternative conceptions. (Note 5) After all, the classic administrative problem, identified by Herbert Simon more than a half-century ago, is the "bounded" rationality of bureaucrats, always limited in their knowledge of the situation, the issues, and the available alternatives (1945). Although this dilemma led Simon to work on artificial intelligence devices that would supplement these limits, it may be that improved and timely access to relevant research can expand the rationality of the policy process. (Note 6)

I would also note, before presenting the policymakers' reflections on online sources of social science research, that this focus on research-for-policy has been perceived in Great Britain as a threat to academic freedom, as well as, in the case of education research, to the professionalism of teachers (Ozga, 2000; Humes and Bryce, 2001). The increased influence of research on government policy also concerns American political scientists Anne Larson Schneider and Helen Ingram (1997), as this increase has done little to counter widespread public disillusionment with the U.S. government since the 1960s. In their view, the growing influence of "scientific and professional perspectives" can further alienate citizens, with the complexity of their approaches and exclusionary jargon. This use of science may seem to increase the rule of reason at the expense of political maneuvering, but its inaccessibility can further shrink the public sphere and public deliberations, as research reduces social issues to technical ones that can be resolved through expert advice and "best practices" (1997, p. 153). (Note 7)

For Schneider and Ingram, science has become a tax-supported "establishment" of limited accountability, which can leave the public wondering why such work needs public support. They identify the National Institutes of Health as an example of scientists securing "overgenerous funding" benefiting "advantaged populations," in their estimation, as well as scientists and drug companies (1997, p. 164). (Note 8) They are also concerned with the considerable number of scientists employed by governments, as this can reinforce the insular power of the bureaucracy. (Note 9) The obvious point of caution is that policy officials, politicians, and interest groups can selectively represent research studies, if not selectively misinterpret the results (Barker and Peters, 1993). For Schneider and Ingram, what the policy process is missing is a democratic commitment to "re-energize people and create an educated, enlightened active citizenship" (1997, p. 7). Their concern harkens back to the democratic divide between John Dewey, who believed in fostering and relying on an increasingly informed public who have the "the ability to judge the bearing of knowledge supplied by others upon common concerns" (1926, p. 209), and Walter Lippmann (1922), who held that a complex democratic state was best governed through "expert mediation" and "organized intelligence." (Note 10)

My own efforts, very much on the side of Dewey, have been to explore ways of improving access to social science research for the benefit of both public and policy officials. The Public Knowledge Project, with which I work, seeks to improve not only the scholarly quality of research, but what I would refer to as its "public quality," in the sense of enhancing its public access and intelligibility. (Note 11) As a foundation for this work, the Project studies the ways in which policymakers, the press, practitioners, parents, and the public, as well as faculty and students, utilize knowledge from online sources. On the basis of this understanding, we are pursuing better designs and structures for organizing, presenting, and integrating research with related resources, such as existing policy initiatives (Willinsky and Wolfson, 2001).

The study presented here focuses on the current levels of engagement with academic research among policy officials. It is intended to serve as a baseline for current usage, as well as to identify a number of key issues that shape the impact that this knowledge has on the development of policy. My argument is that we have an opportunity and responsibility within the social sciences to increase the public and scholarly value of our work. New online publishing technologies, as they increase access, can do more to inform and expand deliberations among people and policymakers. This study asks policymakers about how this new medium is assisting or impeding their use of this research. The resulting interviews suggest a number of steps—including both the labeling and context provided for the research—that would vastly improve the organization, and thus the value, of policy-relevant research online for both public and policymakers. The policy officials in this study were keen to engage with online research, and it appears that the impact on their work would only increase as access and organization are improved. Otherwise, too much of what we know about matters of concern to government policy goes unheard and unseen by those in a position to do something with it, as well as by those who suffer the consequences of this ignorance.

Method

We interviewed 29 civil servants, policy analysts, research officers, and librarians, drawn from the Canadian federal and provincial governments (Table 1). The government departments and divisions that agreed to participate in this research project selected areas for us to conduct interviews in which research might well come into play. The officials (with only two women among the sample) were often referred to us by their superiors, as we explained the nature of our work, as people who were involved in the research and information gathering aspects of their government agency. The Policy Research Initiative, for example, had divisions concerned with Knowledge Integration, as well Social Cohesion and Sustainable Development, suggesting a growing interest in the sort of long-term and broad conceptualizations that the social science research community addresses. There were more prosaic, but nonetheless vital, areas included, such as Employment Insurance and Monetary Analysis.

Table 1
Number of Policy Official Participants by Division and Organization

No.	Division	Government Organization
2	Policy Panning	Agriculture Canada
2	Strategic Policy	Agriculture Canada
3	Monetary Analysis	Bank of Canada
1	Knowledge Transfer	Canadian Health Services Research Foundation
1	Environment Division	Canadian International Development Agency
1	Global Issues Bureau	Dept. of Foreign Affairs and Int'l Trade (Canada)
1	Policy Planning	Dept. of Foreign Affairs and Int'l Trade (Canada)
2	Policy Research Statistics	Department of Justice (Canada)
1	Employment Insurance	Human Resources Development Canada
2	Human Resources Partnerships	Human Resources Development Canada
4	Social Policy Directorate	Human Resources Development Canada
4	Librarians	Legislative Library (Ontario)
1	Knowledge Integration	Policy Research Initiative (Canada)
1	North American Linkages	Policy Research Initiative (Canada)
1	Social Cohesion Group	Policy Research Initiative (Canada)
1	Sustainable Development	Policy Research Initiative (Canada)
29	**Total**	

The questions posed to the policy officials covered three areas of interest, beginning with their "general regard for scholarship and research," including the role it plays in their work and their means of access to it (Appendix 1). The interview went on to consider their use of "electronically available research," including search strategies, favorite sources, and challenges and foreseeable changes. It concluded by asking them to consider two designs for organizing research: one was a prototype site we had previously developed, Policy.ca, and the other a model for providing individual studies with a context. (Note 12) Some subjects declined the offer of anonymity in this research, and their names are used in the analysis that follows, which has been grouped by the nature of the government organization, that is, by government department, agency, research initiative, and library.

Policy Analysts in Government Departments

The mission of Human Resources Development Canada is "to enable Canadians to participate fully in the workplace and the community." (Note 13) In this department, we spoke with an "employment insurance" policy analyst, who described seeking a "diagnositque" from his reading of researchers, stakeholders, trend analyses, and developments in the field, such as a corporation that had a great day-care policy. He was a regular user of government information sites, such as Statistics Canada, but he also found himself often enough facing perplexing Google searches that produced thousands of hits on a topic. Still he admires the speed of information retrieval online, once the wheat has been separated from the vast amounts of chaff: "What we have done is complement the telephone and fax, and what we've got now is much faster." He also made a most interesting case for hard copy by referring to "the cognitive processes necessary for understanding the material" which were assisted by the ability to "spread it out on a desktop to compare it." What also has not changed is that his department still tends to work with the "major players" in the research field, using them to ensure the quality of information as well as funnels and filters for additional research, which may well limit a policy analyst's reach for fresh perspectives. Still, he looks forward to when "more sophisticated [research] portals are introduced" which will provide access to both overviews and full-text studies. He felt that the universities should make an effort to match the Government On-Line (GOL) initiative with a similar approach to research.

In the Human Resource Partnerships Directorate of this same department, we interviewed two senior policy advisors, Silvano Tocchi and Nicholas Wise, who both work with corporations to improve employment opportunities. They, too, rely on an academic network to filter and serve as a conduit for research. They also use it as a "sounding board" for ideas, suggesting a more interactive policy involvement for researchers, if a rather select, handpicked group. Tocchi and Wise see their own work as immersed in social science research, with their own reports and internal research adhering to academic standards. Given the increasing availability of relevant information, including research, they find themselves playing a mediating role with research in their division, or as Wise explains, "sometimes [there is] a need to 'tailor' information in a manner that is more appealing or better understood within the department—packaging it in ways that make it meaningful." They feel it important to browse recent research, without always going after a specific topic, and to that end they subscribe to a number of services that supply a weekly compendium of abstracts or summaries. They were concerned, finally, that efforts to improve information systems would provide links to research carried on outside of the universities, whether to private research companies or to similar departments in other jurisdictions.

With the Social Policy Directorate of Human Resources Development Canada, we held a small seminar with four of its members. This group provides both quantitative and qualitative research support for policy development, especially as a starting point for new initiatives, that also identifies gaps in what is known. Despite the gains through Internet access, they felt that much relevant university research is locked away; they want better systems, greater organization of materials, and more open access to "the real 'meaty' and interesting material" as opposed to available summaries and overviews. Difficulties with getting at this research only meant, they pointed out, that think tank and interest group reports can exert greater influence through their accessibility: "Often you may have only an afternoon to come up with the points you need, so you are going to hit the key groups you know are involved in this area, find the pieces there, and not go any further." Yet they felt that overall the Internet's increased access to research has added to the division's general level of engagement with research: "People are now doing much more of their own research, whereas before there was a select few people or they went to the library and asked a librarian to find information for them. It is now much easier for an individual to do it themselves."

The Department of Justice was the site of an interview with two Research Officers working on issues of cyber-crime. They also followed the traditional pattern of working with a select set of academics in reading their research, consulting with them personally, and inviting them to roundtables. The Internet, however, had enabled them to expand the group they consulted. Still, they shared concerns with others about the overwhelming amount of information on the Web, as well as about the veracity of what is found there: "Although you save time in retrieving the document, you must spend it trying to verify the authenticity of certain material." To help other government officials deal with this issue, they publish "Just Research," which provides a guide to valuable websites for Department of Justice employees. They also expressed a concern that universities do more than just put their work out there: "We need greater ease of connection to places such as university research sites (possibly through a portal or channel), without the frustration of everyone finding the sites always on their own. Then it will be more like going to a traditional library catalogue, where someone is keeping these links up-to-date." This interest in a centralized site included gaining access to studies going back some twenty years, to permit a historical perspective of the very study of justice issues. They were also very articulate about the need for integration among different types of academic literature, from conference papers to meta-analyses, as well as wondering about copyright issues and access to "research data, instruments, and methods" which might be of interest. They were clearly ready for a far more sophisticated approach to online access to research than is currently available.

In the Department of Foreign Affairs and International Trade, we spoke with an International Drug Officer responsible for international comparative analyses of drug policies. The Internet now dominates his information seeking activities. He quickly pointed out that he has no shortage of information on drugs, which only meant that ascertaining "good and credible information" was his biggest challenge. He wondered if having third-party authenticators or guarantors of quality would not resolve this problem. He certainly appreciated the ability to corroborate studies and crosscheck research claims using the Internet, although it was not always possible to do so. The Senior Policy Advisor whom we interviewed from this department also spoke of his increasing dependence on the Web, with up to 80 percent of his research, he estimated, conducted there. He had a real interest in seeing complete studies rather than just abstracts, although he found pay-per-view services for full-texts an economic roadblock. Still, he also reminded us that policies are far more informed by classified information—from government and intelligence sources—as well as by public resources posted by foundations and other organizations, in part because academics seem to lag behind, in his opinion, in placing information on the web. He felt that research needed a context—especially with access to other works by an author—to be judged useful and reliable: "Material and analysis outside of the study itself is useful; and can, perhaps, guide one's own interpretation, for example, a legal decision is much more understandable when the commentaries are reviewed."

In Agriculture Canada, we held two sets of interviews, the first interview with two policy analysts from the Strategic Policy Branch, which is concerned with farm economics. The two posed an interesting contrast to each other. One of them relied on a small set of sites, largely government run, such as StatsCan, and worked with a small circle of researchers: "I usually know the academics that are writing in my area." The other analyst browsed widely, constantly in search of new developments. He believed in the "serendipity" factor, using the example of how he had come across an Australian agricultural initiative that has now led to a collaborative effort between the Canada and Australia. Still, for both of them, the Internet had become their main source; the one could not remember the last time he had been to the departmental library in the basement and the other had been led there by online resources. They were interested in how the Internet could provide some historical depth to their review of research, enabling a review of some thirty years of research. In considering the models for improving the organization of research that we have been developing, they wondered about gaining access to Master's and Ph.D. theses ("which we are seldom aware of") and in being able to establish a context for research—"Reports on reports (reviews), contra views, and summaries are often more revealing than the actual studies." They were also interested in "some form of 'certification,'" as well as knowing who funded the research, as a study "funded 50% by Monsanto does suggest caution and skepticism when reviewing the results."

The other two policy analysts whom we interviewed from Agriculture Canada were from Policy Planning. Here we found real concerns with the failure of research to address the current political issues in agriculture, often forcing them to rely on the work of other policy analysts, whether here or abroad. They were concerned that the resulting absence of information and data has meant that "future policy solutions are more constrained by where we are, where we have been, and our current domestic situation." They would obviously like to have a greater array of perspectives come into play in policy development. As it is, they tend to stay with a small Canadian academic community—maybe 25-30 professors—who are conducting relevant research for their area, noting that while they come across others, it requires checking "what else they have published" and the credibility of their institution—"If Harvard has hired them...." They were also sensitive to how the use of research in policy settings invoked issues of jargon and translation—"translating academic work into a language our executives understand is my job." Yet there are areas were they feel well served by the Web for developing their professional skills: "I use the Internet more for locating management and strategic thinking material, than policy material, i.e., Kennedy School of Government, Harvard, Stanford, or universities with good management schools, I know who the significant management writers are, and then see what they have made available on the site."

Among the policy officials working in their four government departments, the Internet has taken hold as a primary information source, offering ease and speed, while posing challenges of extracting value and establishing credibility. There is clearly room for social science research to play a greater part in the policy process, for expanding the range of ideas considered and opening up the government's thinking about possibilities that extend beyond the closed traditional circles of academics.

Policy Analysts in Government Agencies

We worked with two government agencies, the Bank of Canada and the Canadian International Development Agency that are very involved in policy issues, if at one remove from the legislative forum of Parliament. At the Bank of Canada, we spoke with three men involved in monetary modeling, applied studies, and policy research. In each case, they told us, research is what they do. They are economists who both produce reports for the Bank and publish research in academic journals. They pointed out that economics is one academic discipline that has a central open-access repository, RePEc (Research Papers in Economics) that provides free and centralized access to an enormous number of articles, reports and working papers. They see RePEc as a real advantage to their work as researchers, just as they appreciate having electronic access to complete sets of economic journals online, running back to the turn of the century, through J-STOR. (Note 14) Still, they sometimes feel frustrated by disciplinary boundaries that separate other relevant areas, such as biology, mathematics, and statistics, from this bounty of economic literature. And while they cannot imagine working now without the Internet, they would like to see greater consolidation of access and indexing, as well as more full-text resources and citation linking, as the search for research resources still seems unnecessarily difficult. Their opinions speak to the technology's ability to foster a greater demand for its powers than it can, at any given moment, meet.

Caroline Caza is a Senior Environmental Policy Advisor, with the Canadian International Development Agency, which is devoted to supporting "sustainable development activities in order to reduce poverty and to contribute to a more secure, equitable and prosperous world." (Note 15) Although she does not see herself as a regular consumer of research, Caza's impression is that academic research does influence the formation of policy, observing that academics are hired on occasion to review the literature and present feasible policy perspectives. She pointed out that within her own agency the press was on to create a "knowledge-based organization," which would increase its capacity to create effective policies. In her own work, Caza prefers reviews of current research which she finds "much more useful forms of information," while individual studies can be "too long" and "too narrow to be able to synthesize the diversity of opinion you that need in order to make a credible policy decision." Yet she also felt that what was frustrating about some information rich websites was "the lack of depth and breadth in, for example, Policy.ca" (referring to the Public Knowledge Project site). And while senior policy officials are always asking for quality information, they may not be in a position to judge that quality themselves. For her part, she has let her subscriptions to academic journals, acquired during her university days, lapse as they failed to serve in her work: "They didn't allow me to be credible by bringing new scientific information to a forum in a way I could use effectively." The exception was Conversation Biology which Caza continued to receive for its Letters to the Editors, which she felt presented leading-edge issues and any controversy surrounding them. So while she was turning from print sources, she was still seeking the ideal information websites which offered authentic and credible information, in what was an all too common theme among policy officials in our sample.

Government Research Initiatives

We spoke with representatives from two of the Government of Canada's research agencies: the Policy Research Initiative and the Canadian Health Services Research Foundation. The Policy Research Initiative was established "to strengthen the federal government's policy research capacity" and sees itself as "a catalyst for the development of *knowledge, people and community*," according to its website (original emphasis). (Note 16) Here, the government is fully engaged in bringing research to bear on areas of Canadian policy. It publishes *Isuma: Canadian Journal of Policy Research*, which features an editorial board made up of academics (although one PRI representative we spoke with felt it was "not as in-depth as a traditional academic journal"). The journal offers an open-access electronic version as well as a print run that is distributed "free to a large cross section of Canada's policy research community."

The Policy Research Initiative also publishes *Horizons*, "a snapshot of emerging, cross-cutting research in the Canadian and international policy environments." It also has a Policy Research Data Group that works with other departments facilitating "the development of databases required to carry out research in priority horizontal policy areas."

Within the Policy Research Initiative is the Knowledge Integration Project which is directly involved in exploring new approaches to managing and accessing scholarly research, including the use of emerging standards for document management systems, such as Open Archives Initiative, Government Information Locator Service, and Dublin Core Metadata Initiative. The representative of the Knowledge Integration Project with whom we spoke saw his role within the policy community as developing strategies for the long-term acquisition and storage of "knowledge objects," reflecting a government commitment to using new technologies to increase the play of research and other sources of information in the policy process. His concern that it was often difficult to locate and work with the data sets behind major research studies suggested the need for developing far more comprehensive scholarly communication systems that would enable independent and critical analysis of research, as well as greater collaboration and reanalysis of data. As an experienced research reader, he was less than happy with how research was currently made available, both economically and structurally: "Until we define or restructure the 'Elsevier' economic business model for publishing, we are going to be stuck with very high transaction costs for finding information.... We need both a common format standard and a common conceptual standard." The inhibiting factor, in his eyes, is "the transaction cost of searching on the Internet" which he finds "phenomenal." And while he tended to use materials that were only one or two years old, he also saw a place for keeping a "500-year time frame" in addressing major policy issues, pointing to an interest in both historical research, as well as research from earlier periods, as did others in our sample. Finally, this knowledge integration officer expressed a strong interest in a quality indicator or systems of endorsement that went beyond basic peer review, of the sort perhaps, used in bibliometrics to evaluate the "impact factor" of a journal (by how often its articles are cited).

A second representative of the Policy Research Initiative, this time from the North American Linkages division, was very clear about the foundation that research provides for policy. She felt that it did fall to the government to strengthen scholarly research in areas of particular policy interest. The government could then draw on these researchers to build an advisory group who could be counted on to keep the government informed and ensure the quality of the resulting policies. In terms of her reliance on the Internet as a source of research, she pointed out, as did others, that assessing the quality of the research is as much an issue as accessing it: "You have to exercise a considerable amount of independent judgment about the veracity of what you are reading." She believes that researchers in her divisions are increasingly relying on the Internet—where the quality of information is still difficult to ascertain—when they should be seeking out the known quantities in print sources. Her worry is that the Internet leaves readers with a misleading sense of having consulted, or at least browsed, all that there is to know in a particular field. Otherwise, it could mean that the Internet, at this point at least, may represent a shift in the focus of attention on research—to what is readily available online—with no real expansion in the consideration of what is known and no guarantee as to the quality of that knowledge.

The Policy Research Initiative representative from the Sustainable Development group, Paul Halucha, offered a glimpse of a still very powerful method of working directly with a small number of researchers to directly address the issue at hand. He described how he brought together academics to identify eight relevant trends in research on sustainability which resulted in an "interdisciplinary knowledge statement." He saw this process offering a "single window" into the government which researchers could use to bring their work to bear on policy. Yet it was exactly that, a single window, and not an open door. It enabled direct engagement between scholar and policy official, but it was more of a closed process, and needs to be supplemented at least, by more open processes. It was, to be fair, his way of dealing directly with the knowledge demands of his work: "Government officials are faced continually with the daily pressures of producing knowledge." He felt the need for better ways of managing the deluge, as "information...is piling up...and there comes a point at which one gets into declining returns." He also worried about how the Internet does not afford access to older materials, which would diminish "one of the classic public policy values" of "deliberation, rational deliberation." The challenge, then, is to devise systems that do not leave one reverting to narrower, traditional processes of consultation because the new is simply too much, too quickly.

A third representative of the Policy Research Initiative with whom we spoke was Michael Mackinnon, from the Social Cohesion Team. Mackinnon spoke of how the concept of "social cohesion"—which is concerned with the degree of people's economic, social and political participation in a society—was derived from European sociological research, which had been recently popularized on this continent by Robert Putnam's work (2001). Mackinnon spoke of the need for better organized websites with taxonomic structures and tagged materials, as well as for research written in comprehensible language, with some way of verifying its status. He also pointed to the value of portal sites that served as well organized gateways to a wide range of resources, commending Policy.ca, as well as Canadian Social Research Links, which is run as a hobby by Gilles Séguin, a Human Resources Development Canada employee. (Note 17) While often satisfied with being able to read research abstracts, Mackinnon also showed an interest in having the opportunity to go deep and review the actual data sets behind studies, as well as in being able to consult summaries, overviews and syntheses. While all of this would be "of significant interest to policy analysts," he wondered who could afford to produce this level of research support, as an add-on to existing systems, while what may be needed is a new system in which summaries, portals, and taxonomies are all part of the fundamental design of scholarly communication.

The Canadian Health Services Research Foundation is a non-profit organization endowed by the federal government to support both basic and practitioner research on management and policy issues in the area of health services and nursing. (Note 18) It also seeks to bring together decision-makers, policymakers, practitioners and researchers to facilitate the dissemination of this research. We spoke with Michelle Campbell, Assistant Director, Knowledge

Transfer Division, who works to ensure that the available knowledge gets to the policy researchers, decision-makers, and administrators in Canada's health systems. As part of this job, her group does translate research on occasion into common language, on the one hand, and has hired a "knowledge broker," on the other hand, to work with academics. She has learned through this work that while researchers feel they can draw no conclusions when they have only 20 per cent of the answer, policymakers feel that 20 percent of the answer is about as good as it gets. Given than, policymakers are often frustrated by the researchers' reluctance to make recommendations on the basis of the existing evidence.

Campbell felt that the research abstract should clearly present the implications of the study, for at that point, the policymaker decides whether to examine the study's conclusion, while only occasionally going back to the findings and discussion. That said, she felt while it was important for policymakers to have access to complete studies, she also wanted a way of moving up from a study to a synthesis or meta-view of the larger research issues, given the range of pertinent research: "The actual breadth of sources is huge...and no one can check them all or find one place where an individual can check them all. Secondarily, the type of research is much more complex to do than traditional clinical intervention research, so that quality is much trickier of an issue." And this time, when the quality assurance issue of information on the Web came up, it was in reference to the "gray literature" which included relevant but unpublished reports and commentaries. And while Campbell would like to see more warranted research online, the efforts of her own agency are sometimes curtailed in this regard. The Canadian Health Services Research Foundation has to be careful in its handling of the research it has sponsored so that it did not diminish that research's ultimate "publishibility." This could well result in limited public access to publicly sponsored research. The research incentive system of publish-or-perish was working against the very mission of the Foundation. There should be a way to provide the traditional quality control of peer review while providing wider access to this publicly sponsored work on improving health care.

Among the Librarians

To complete the picture of how policy officials work with research, we visited the Ontario Legislative Assembly Library and spoke with two librarians and two research officers. Their job, as one of them put it, is "to help members [of the Legislative Assembly] improve the quality of debate." That is, the information they provided came into play after the policies had been drafted and were being considered by the Assembly. "We're not often asked to help at the inchoate stages of policy," one of the research officers told us. They saw the information serving as a background to the work of the Assembly, whether in explaining the policy to opposition party members or to the press. Using research to inform the discussion of policy, which was not raised by the policy analysts we interviewed, clearly plays a critical role in a democratic society for setting the scope and terms of the debate. One of the research officers works extensively with academic research ("I have to keep up with political science research," he said. "I need the journals at my fingertips") while the other research officer, who works with legal matters, deals more often with the legislation itself, mainly in comparisons of statues across jurisdictions, for example. For the former, greater online access to well-organized bodies of research would be welcome, especially against the time constraints of having to prepare reports quickly that are supported by current research. And the currency of research by the time it reaches print, he noted, was often inadequate. The acquisitions manager of the library pointed out that increasing subscription costs have made it difficult for the library to retain academic serials, and that online systems were currently being offered as bundled services that required purchasing access to more than was felt warranted: "I don't want to buy the whole thing if I need only 10 per cent of it." It suggested that another sort of system was required to otherwise prevent a decline in the consultation of academic sources of information on policy matters.

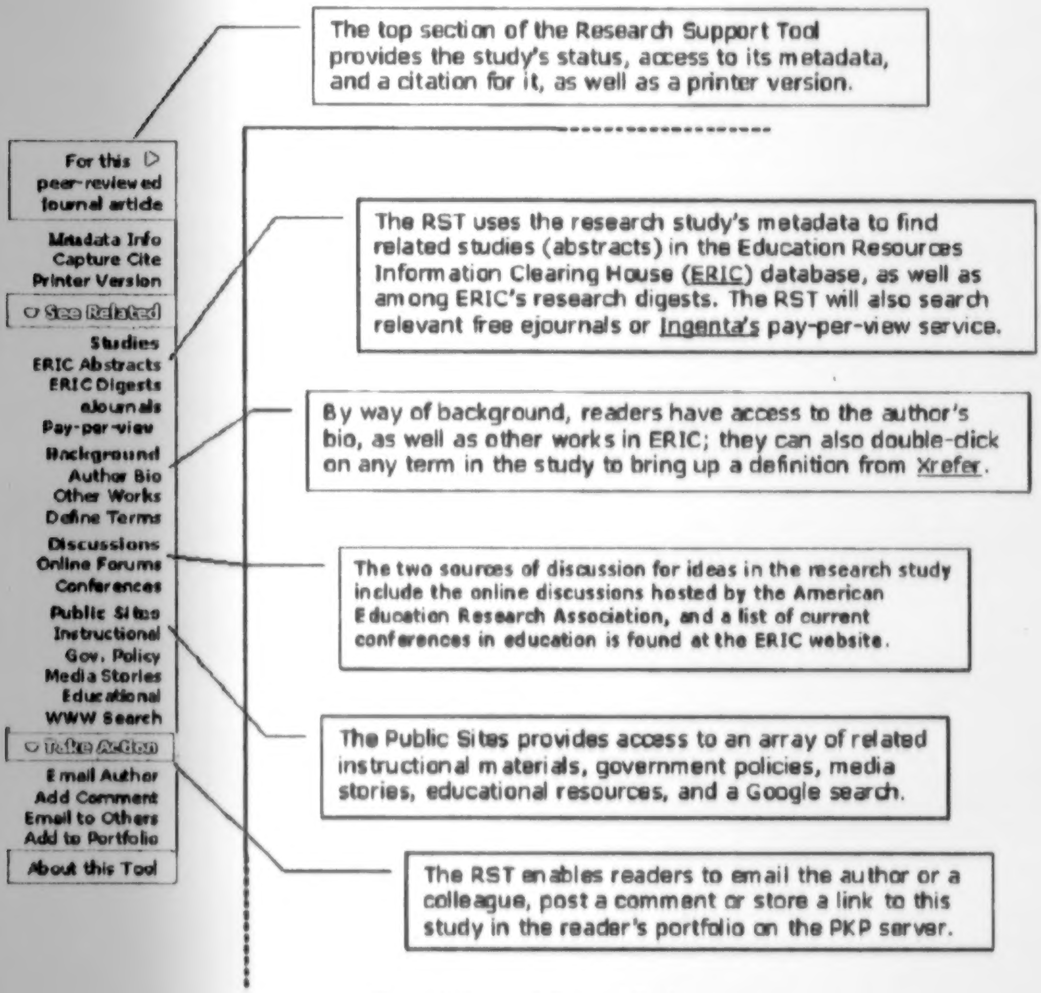
Discussion

The interviews with these Canadian policy officials and related personnel make it clear that the Internet is now a favored source of information within government. It is used to tap into the research that is consulted as part of the policy process. It is also clear that the research that is most easily accessible, through portals and open-access sites, is most often consulted, as policymakers referred to how readily they were dissuaded from using pay-per-view and subscription services in their pursuit of knowledge. This means that they are tapping into a skewed and somewhat haphazard view of the current state of knowledge on a given topic. This could be seen as a further argument for establishing an "open access" economy of scholarly communication that would make scholarship freely available online. (Note 19)

Whatever form online scholarly publishing systems finally take, nearly everyone we spoke with agreed on the need for a system that warranted the status or credibility of the research, from clearly marking it as "peer reviewed" (with an accompanying explanation of the process) to providing citation statistics, which indicated whether others have referred to the paper, and in what context. (Note 20) Yet there is more to research online than peer review. The Internet also offers pre-prints and working papers referring to research, which has yet to be reviewed, but these are clearly marked with at least one set featuring a "warning" for "causal readers." (Note 21)

Among the other suggestions I would draw from the participants comments is the need for non-academics to establish a coherent context for a given study that enables readers (1) to review other studies by the author, as well as related studies both current and those going back decades, (2) to go deep into the body of the study, as far perhaps as the research instruments and raw data set where possible, (3) to rise up, above the specific study, to an overview or summary of the larger field of inquiry, (4) to consult related resources, such as relevant court decisions, as one policy maker mentioned, or related policy in other jurisdictions, (5) to access glossaries or reference materials that further open up the language and ideas to a wider readership, and (6) to interact with researchers, to comment on their work and consult directly with them, providing the researcher with a sense of that wider audience which will, in turn, give shape to how they express their findings.

Each of these abilities to work with research is certainly feasible in an online environment. For example, on giving some historical depth to the ability to locate related studies, the J-STOR project has been putting complete sets of journals, dating back to the nineteenth-century online. (Note 22) For our part, we drawn on these interviews to develop a "research support tool" that would provide a way for readers to readily move from an individual research study to related resources not only in research but in policy, practices, the media and other areas (Figure 1). We are also working on improving the indexing of online research to enable more precise and accurate searching for academic resources. The next phase of the larger study is to test the viability of such a support system with the participants of this study, the majority of whom expressed a willingness to take this next step.



The top section of the Research Support Tool provides the study's status, access to its metadata, and a citation for it, as well as a printer version.

The RST uses the research study's metadata to find related studies (abstracts) in the Education Resources Information Clearing House (ERIC) database, as well as among ERIC's research digests. The RST will also search relevant free ejournals or Ingenta's pay-per-view service.

By way of background, readers have access to the author's bio, as well as other works in ERIC; they can also double-click on any term in the study to bring up a definition from Xrefer.

The two sources of discussion for ideas in the research study include the online discussions hosted by the American Education Research Association, and a list of current conferences in education is found at the ERIC website.

The Public Sites provides access to an array of related instructional materials, government policies, media stories, educational resources, and a Google search.

The RST enables readers to email the author or a colleague, post a comment or store a link to this study in the reader's portfolio on the PKP server.

Figure 1. Research Support Tool

Figure 1 represents an initial design for a tool that would provide studies in the field of education with a greater context for interpretation, judgment, and utilization by policymakers (and the public), based on the suggestions made in the interviews in this study. A working demo of this tool can be found at http://pkp.ubc.ca/demos/rsttour/index.html.

As encouraging as this potential for improving policy officials' greater engagement with research may be, there remains the political caution about research's influence expressed by Anne Larson Schneider and Helen Ingram, whose work I discussed above: "Where scientists and professionals have the greatest influence, however, the consequences are damaging to democratic values" (1997, p. 181). The increasing presence of research, through online technologies may well make citizens feel disempowered and "less capable of self-government" (p. 185) in the face of what Walter Lippmann identified many years ago as the "organized intelligence" necessary to run the modern state (1922). (Note 23) The policy officials with whom we spoke did lend credence to Schneider and Ingram's fear "that scientific and professional networks...have colonized government so that there are no public arenas but only bureaucratic maneuvering among privileged specialists" (1997, p. 188). Not only did we speak with government economists, for example, shaping policy, but more than one policy official referred to working with a select group of academics who served as research filters and policy consultants. The danger of this limited consultation, outside of public arenas, can be challenged, we believe, by improving the online presence of research as part of knowledge's public sphere. Creating open access to research would help policymakers, as well as the public, consult a much broader and more diverse range of information. With improved indexing, policymakers and the public would be able to locate innovative and fresh perspectives on specific and pressing issues, as well as ensure that they could consult contrary opinions and gain a sense of the range of approaches and opinions on a topic.

What is clear, from a democratic perspective, is that improving policymakers access to research has to be balanced by similar improvements in public access. This represents the democratic check on research's political impact factor, but it also promises to raise the level of public deliberation and the very quality of democratic processes. Improving both policy and public access to research entails many of the same concerns with credibility, overview, integration, and context. In the United States, for example, science's political contribution has been officially channeled through advisory bodies, such as the National Academies (with the Royal Society of Canada proposing a similar approach for the Canadian government). The Academies' reports are made public and typically try to represent a coherent consensus among experts, with due consideration given to policy implications. As reassuring as such reports can be to the public, Stephen Hilgarter's investigations have revealed the degree to which they are shaped by "stage management and struggles over the enclosure and disclosure of information" (2000, p. 146). These carefully reviewed reports may still misrepresent the state of dissent among researchers, just as the prestige and authority of these research reports can seem to squelch the deliberative and difficult public processes that make for enlightening, if sometimes exasperating, democratic experiences.

The democratic impact of research is likely to be advanced, given what this sample of policy officials revealed in their interviews, by creating a far more open, coherent, and integrated access to scholarship in ways that might serve policymakers and the public alike. The greater coherence and integration of scholarly communication that might be achieved through these new technologies should not be directed at overwhelming policy processes with the dictates of a scientific consensus. Rather, these new systems should be concerned with bringing the rich diversity of values and findings of research into public view and play. These systems need to make research more fully a part of a public sphere, marked by the constant contest of ideas, in the search for greater understanding of what is and can be known.

As social science researchers, it will not do, I contend, to imagine that our work can stay safely removed from the messy world in which people live and govern. We need to see our work as part of that life. In a few short years, the Internet has already increased the presence of research in the policy process. It has to a large extent replaced the traditional print sources, and well before this new communication medium has arrived at anything like a sufficiently well organized way of handling the publication of research. It falls to researchers, then, to take hold of the democratic possibilities, and to be guided by the clear cautions, in developing new systems for scholarly communication for both public and scholarly use.

Acknowledgment

I wish to thank Michael J. D. Sutton for his talented assistance with the interviews in this article, Anne White for her attentive reading of it, the Max Bell Foundation for its support of this work, and the editor of this journal for his fine editorial assistance.

Notes

1. See Alexander and Pal (1998), Hague and Loader (1999), Heeks (1999), and Wilhem (2000). The potential influence of this new technology on citizen consultation, for example, has only begun to be experimented with, and with mixed results. For our experiments with public and professional consultation on education policy, see the Public Knowledge Policy Forum under prototype websites of the Public Knowledge Project (http://pkp.ubc.ca) and Klinger (2001).
2. The American Educational Research Association, for example, lists roughly100 e-journals that offer free, full-text access, including the well-respected *Educational Researcher* and *Teachers College Record* (http://aera-cr.ed.asu.edu/links.html). RePEc: Research Papers in Economics provides open access to 102,000 working papers and 56,000 journal articles (http://repec.org/). The University of Pennsylvania provides an excellent guide to electronic journals in all fields (http://www.library.upenn.edu/webbin5/resources/ejspublic5.cgi).
3. On the recent new education bill, see Schemo (2002). Also Education policy's stake in research, at the U.S. federal level, is reflected in the U.S. Education Department's 2001 fiscal allocation of $185 million for education research and $80 for education statistics. Though it pales before the National Institutes of Health's $20.3 billion allocation or the Defense Department's $1.4 billion for "basic research," it represents a considerable and stable investment in creating knowledge that will presumably serve education policy and practice, and President Bush has proposed a 28 percent increase for educational research 2003 (Burd and Southwick, 2002).
4. On the challenges faced in the United States over evidence-based education, see Loveless et al. (1999), and on its pertinence for education (Author, 2001).
5. See Friedman for a comparison of in-house bureaucracy, private consultants, and academic sources for policy analysis, with the academic "loner" judged to be often out of touch with "shifting sands of policy terrain," while their advice possesses a "somewhat abstract air" and they are "rarely available" when needed (1987, p. 162).
6. It may also be worth noting, on the other hand, the "social functions of ignorance" among bureaucracies, which Moore and Tumin argued, preserves traditional values and fair competition, as well as privilege and stereotypes, giving it a certain value for policy officials over the headlong pursuit of knowledge (1964).
7. Also see Stephen Hilgartner, who has observed that "governments find expert advice to be an indispensable resource for formulating and justifying policy and, more subtly, for removing some issues from the political domain by transforming them into technical questions" (2000, p. 146).

8. The National Institute of Child Health and Human Development have had considerable influence on education policy, although not without controversy among researchers: see Allington and Woodside-Jiron (1999), with response from Mathes and Torgesen (2000); and see Taylor et. al. (2000) with response from Foorman et al. (2000).
9. Here Schneider and Ingram (1997, p. 154) cite Max Weber's caution in his essay, "Three Types of Legitimate Domination" (1978, pp. 991): "Bureaucracy naturally prefers a poorly informed, and hence powerless, parliament at least insofar as this ignorance is compatible with the bureaucracy's own interest."
10. Lippmann on the development of democracy in America: "The more enlightened directing minds have called in experts who were trained, or trained themselves, to make parts of this Great Society intelligible to those who manage it" (1922, p. 234). See Willinsky (2000, pp. 37-42, 100-105) and Aikens (1999).
11. See Willlinsky (2000b) as well as the Public Knowledge Project (http://pkp.ubc.ca).
12. See the Public Knowledge Project website (fn. 10) for access to the two designs.
13. Human Resources Development Canada (http://www.hrdc-drhc.gc.ca).
14. RePEc has some 160,000 items, including 60,000 online (http://ideas.uqam.ca); Also see, J-STOR: The Scholarly Journal Archive (http://www.jstor.org/)
15. Canadian International Development Agency (http://www.acdi-cida.gc.ca/).
16. Policy Research Initiative (http://policyresearch.schoolnet.ca).
17. Policy.ca (http://policy.ca); Canadian Social Research Net (http://www.canadiansocialresearch.net/).
18. Canadian Health Services Research Foundation (http://www.chsrf.ca/).
19. See Public Library of Science (http://www.publiclibraryofscience.org/) and Budapest Open Access Initiative (http://www.soros.org/openaccess/).
20. See the NEC ResearchIndex for an online and open access system for representing the actual context in which a study has been cited by another study (http://citeseer.nj.nec.com/cs).
21. See NetPrints: Clinical Medicine and Health Research set up by the British Medical Journal and Stanford University Self-Archiving Initiative which includes this statement boldly set up at the entrance to the site: "Warning: Articles posted on this site have not yet been accepted for publication by a peer reviewed journal. They are presented here mainly for the benefit of fellow researchers. Casual readers should not act on their findings, and journalists should be wary of reporting them" (http://clinmed.netprints.org/).
22. J-STOR (http://www.jstor.org/).
23. G. Scott Aikens sees the resolution of the Walter Lippmann and John Dewey struggle between expert and public as lying in the cyberspace potential of public forums and open systems: "Whist Deweyan systems accept the need for rich systems of organized intelligence in complex societies, these can become richer through the active engagement of experts in open and free decision-making systems at the local, regional, and global level" (1999, p. 192).

References

Aikens, G. S. (1999). Deweyan systems in the Information Age. In B. N. Hague & B. D. Loader (Eds.), *Digital democracy: Discourse and decision making in the Information Age* (pp. 179-194). New York: Routledge.

Alexander, C. J. & Pal, L. A. (Eds.). (1998). *Digital democracy: Policy and politics in the wired world*. Toronto ON: Oxford University Press.

Allington, R. & Woodside-Jiron, H. (1999). The politics of literacy teaching: How "research" shaped educational policy. *Educational Researcher, 28*(8), 4-13.

Barker, A. & Peters, B. G. (1993). Introduction: Science politics and government. In A. Barker & B. G. Peters (Eds.), *The politics of expert advice: Creating, using and manipulating scientific knowledge for public policy*. Pittsburgh PA: University of Pittsburgh Press.

Burd, S. & Southwick, R. (2002, January 11). Congress approves increases of $250 in top Pell Grant and $3-billion for the NIH. *Chronicle of Higher Education*, A32.

Dewey, J. (1926). *The public and its problems*. New York: Holt.

Foorman, B. R., Fletcher, J. M., Francis, D. J., & Schatschneider, C. (2000). Response: Misrepresentation of research by other researchers. *Educational Researcher 29*(6), 27-37.

Friedman, J. (1987). *Planning in the public domain: From policy to action*. Princeton NJ: Princeton University Press.

Hague, B. N. & Loader, B. D., (Eds.). (1999). *Digital democracy: Discourse and decision making in the information age*. London: Routledge.

Heek, R. (Ed.). (1999). *Reinventing government in the information age: International practices in IT enabled public sector reform*. London: Routledge.

Hillage, J., Pearson, R., Anderson, A. & Tamkin, P. (1998). *Excellence in research on schools*. Department for Education and Employment, London.

Hillgartner, S. (2000). *Science on stage: Expert advice as public drama.* Stanford CA: Stanford University Press.

Humes, W. & Bryce, T (2001). Scholarship, research, and the evidential basis of policy development in education. *British Journal of Educational Studies*, *49*(3) 329-352.

Klinger, S. (2001) *"Are they talking yet?" Online Discourse as Political Action in an Education Policy Forum.* Unpublished dissertation, University of British Columbia.

Lippmann, W. (1922). *Public opinion*. New York: Free Press.

Loveless, T., Peterson, P., Boruch, R., Cook, T., Gueron, J., Hyatt, H., & Mosteller, F. (1999). *Can we make education policy on the basis of evidence? What constitutes high quality educational research and how can it be incorporated into policymaking*? A Brookings Press Forum. Washington DC: The Brookings Institution.

Mathes, P. G. & Torgesen, J. K. (2000). A response to Allington and Woodside-Jiron. *Educational Researcher*, *29*(6), 4-14.

Moore, W. E. & Tumin, M. M. (1964). Some social functions of ignorance. In B. Rosenberg, I. Gerver, & F. W. Howton (Eds.), *Mass society in crisis: Social problems and social pathology*. New York: Macmillan.

>Ozga, J. (2000). *Policy research in educational settings*. Buckingham, UK: Open University Press.

Putman, R. D. (2001). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Schemo, D. J. (2002, January 9). Education bill urges new emphasis on phonics as method for teaching reading, *New York Times*, A7.

Schneider, A. L. & Ingram, H. (1997). *Policy design for democracy*. Lawrence KA: University Press of Kansas.

Simon, H. (1945). *Administrative Behavior*. New York: Free Press.

Taylor, B. M., Anderson, R. C., Au, K. H., & Raphael, T. E. (2000). Discretion in the Translation of Research to Policy: A Case from Beginning Reading. *Educational Researcher*, *29*(6), 16-26.

Weber, M. (1978). Three types of legitimate domination. In C. Roth & C. Wittch (Eds.), *Economy and society: An outline of interpretive sociology*. Berkeley, CA: University of California Press.

Weiss, C. (1998). *Interview by Miranda Christou*. Cambridge MA: Harvard University School of Education. Available at http://www.gse.harvard.edu/~oie/weiss.html.

Weiss, C. (1991). Policy research: data, ideas, or arguments. In P. Wagner, C. H. Weiss, B. Wittrock & H. Wollman (Eds.), *Social sciences and modern states: National experiences and theoretical crossroads* (pp. 307-332). Cambridge, UK: Cambridge University Press.

Wilhelm, A. G. (2000). *Democracy in a digital age: Challenges to political life in cyberspace*. New York: Routledge.

Willinsky. J. (2000a). *If only we knew: Increasing the public value of social science research*. New York: Routledge.

Willinsky. J. (2000b). Proposing a Knowledge Exchange Model for scholarly publishing. *Current Issues in Education*, *3*(6). Available: http://cie.ed.asu.edu/volume3/number6/.

Willinsky. J. & Wolfson, L. (2001). The indexing of scholarly journals: A tipping point for publishing reform? *Journal of Electronic Publishing*, *7*(2). Available at http://www.press.umich.edu/jep/.

About the Author

John Willinsky
Department of Language and Literacy Education
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
john.willinsky@ubc.ca

John Willinsky is the Pacific Press Professor of Literacy and Technology at the University of British Columbia where he directs the Public Knowledge Project. He is the author of a number of books, including *Learning to Divide the*

World, which won two outstanding books awards, and *If Only They Knew: Increasing the Public Value of the Social Science Research.*

Appendix

Interview Schedule Used With Policymakers

1. Background
1.1 Introduce UBC PKP Project
1.2 Describe your organization, its mission, and your role within it
1.3 What role does your organization play in policy making?
2. General regard for scholarly research
2.1 What role does scholarly (university-produced) research currently play in your policy decisions?
2.2 What is the most frustrating thing you find about accessing scholarly research?
2.2.1 Is there a problem with interpreting the language of the material?
2.2.2 Is there a problem with searching/locating material?
2.2.3 Is there a problem with Internet connection/firewalls?
2.2.4 Is there a problem with the ability to access/capture/store the information?
2.2.4 Is it a problem with intellectual property rights/copyright?
2.3 What possible change could make that frustration go away?
3. Electronically Available Research
3.1 Describe your organization's use of electronic information sources?
3.1.1 Describe your own personal use
3.1.2 What are your leading sources of online information?
3.2 How has your use/approach changed in recent years?
3.3 How do you foresee such use changing in the coming years?
3.4 What are the promising aspects of increased access and interactivity?
3.5 Have you experienced any important obstacles or issues surrounding:
3.5.1 Comprehensiveness
3.5.1.1 How difficult is it to locate and continue to work with 1 or 2 reliable starting points?
3.5.1.2 Is access to full text versions of research material mandatory/desirable?
3.5.1.3 How much depth/richness must a site have in order for you to return to it? How do you determine this?
3.5.1.4 Is detailed, current contact information a requirement for work featured on a site?
3.5.1.4.1 By institution?
3.5.1.4.2 By individual?
3.5.1.5 When do you know you have covered a topic in enough depth to proceed with policy analysis?
3.5.1.6 Is a breakdown by jurisdiction important?
3.5.1.6.1 Local
3.5.1.6.2 Regional
3.5.1.6.3 Provincial
3.5.1.6.4 Federal
3.5.1.6.5 International?
3.5.2 Accessibility
3.5.2.1 Does source language of material provide any constraint?
3.5.2.2 Does the academic language of the research discourage interpretation?
3.5.2.3 Are hardcopy or digital indices used to locate relevant research?
3.5.2.4 Is it more common to want material about a "study" rather than the "study" itself?
3.5.2.5 Is a hardcopy mandatory/desirable, or would a digital copy be sufficient?
3.5.2.6 Does the digital format of research material inhibit your work?
3.5.2.6.1 Adobe pdf
3.5.2.6.2 MS-Word
3.5.2.6.3 Postscript
3.5.2.6.4 HTML?
3.5.2.7 Do you experience any connectivity constraints, (dial-up vs. HS lines, lack of connection)?
3.5.2.8 How would you rate your on-line search skills?
3.5.2.8.1 Novice
3.5.2.8.2 Experienced
3.5.2.8.3 Master?
3.5.2.9 Do you use any current awareness services (i.e., email alerts, etc.)?
3.5.2.10 Any intellectual property problems encountered around copyright & digital rights?
3.5.3 Currency
3.5.3.1 Is the availability of current material mandatory/desirable?
3.5.3.2 Generally speaking, how current is material you presently use?
3.5.3.2.1 <= 1 month old
3.5.3.2.2 <= 3 months old
3.5.3.2.3 <= 6 months old
3.5.3.2.4 <=1 year old
3.5.3.2.5 > 1 year old
3.5.3.3 How old should historical or archival material be in order to make it relevant to your work?
3.5.4 Reliability
3.5.4.1 How do you verify the credibility of both the author and the content of research material?
3.5.4.2 Would direct/immediate contact with journal reviewers be important for reliability?
3.5.4.3 Do you believe that research which is available on-line is inherently inferior to hardcopy?

4. The Knowledge Exchange Model

- 4.1 Introduce the Knowledge Exchange Model
  - *4.1.1 Walkthrough Model*
  - *4.1.2 Walkthrough http://www.Policy.ca*
    - 4.1.2.1 Are there any suggested changes you could make for http://www.Policy.ca?
- 4.2 Given what you know about the Knowledge Cube,
  - *4.2.1 Do you think it could help you access scholarly research?*
  - *4.2.2 Why or why not?*
- 4.3 Imagine for a moment that all obstacles to implementing such a model had somehow been overcome.
  - *4.3.1 How would the model be different?*
  - *4.3.2 What would be its general characteristics?*
- 4.4 Would you be willing to review the final design(s) and the resulting prototype(s)?

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)

Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@airbrasil.org.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

This article has been retrieved 8649 times since January 15, 2003

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Education Policy Analysis Archives

Volume 11 Number 3 **January 15, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article if EPAA is credited and copies are not sold. EPAA is a project of the Education Policy Studies Laboratory.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Limits of Sanctions in Low-Performing Schools: A Study of Maryland and Kentucky Schools on Probation

Heinrich Mintrop
University of California, Los Angeles

Citation: Mintrop, H. (2003, January 15). The limits of sanctions in low-performing schools: A study of Maryland and Kentucky schools on probation, *Education Policy Analysis Archives*, *11*(3). Retrieved [date] from http://epaa.asu.edu/epaa/v11n3.html.

Abstract

The article reports on a study of 11 schools that were labeled as low-performing by the state accountability systems of Maryland and Kentucky, nationally known for complex performance-based assessments. The study shows that putting schools on probation only weakly motivated teachers because the assessments were largely perceived as unfair, invalid, and unrealistic. Administrators responded with control strategies that rigidified organizations, forestalling dialog and learning processes. Instructional reform developed only feebly. On the other hand, some schools remedied inefficiencies and were able to "harvest the low-hanging fruit." The schools struggled with severe problems of teacher commitment.

The proliferation of high-stakes accountability systems in the United States has fast created a new category of schools identified with various labels: Schools on Probation, Schools under Reconstitution, Schools in Decline or in Crisis, Schools under Review, Immediate Intervention Schools, Schools Eligible for Assistance, and so on. Each accountability system has created its own nomenclature, but the underlying structure is the same: Based on a small set of numerical performance indicators, accountability systems identify putative underperformers that are given a limited period of time to reverse growth deficits or decline and that are threatened with more severe penalties upon failure to do so. In the public debate, these schools are also known summarily as "failing."

As of the year 2001, 27 states have had school accountability systems that identify low-performing schools and 14 states stipulate more severe penalties when an underperforming school fails to improve (Boser, 2001). In 1997 the city of Chicago alone identified a hundred or so public schools on probation that managed to have fewer than fifteen percent of their students read at the national norm, as measured by the Iowa Test of Basic Skills (Chicago Public Schools, 1997). To date, the small state of Maryland has identified 200 or so schools statewide. Between 1999 and 2001 alone, the large state of California identified 1290 persistently low-performing schools that are enrolled in the state's "Immediate Intervention/ Underperforming Schools Program." Implementation of the new federal Title I legislation may add further impetus to the phenomenon. What is more, these schools are not randomly or evenly distributed across the states, but in many instances are clustered in districts that traditionally serve poor and disadvantaged minority populations. For example, in Maryland, almost all identified schools are located in two districts; in California, 54 of the 1000 or so school districts with more than 10 schools have at least a third of their schools eligible for California's Immediate Intervention/ Underperforming Schools Program.

Leaving aside the possibility of a mere symbolic purpose of "high stakes," these policies bank on the motivational power of sanctions, currently conceptualized in two ways. In one version, popular among economists, high-stakes accountability is resource-neutral; that is, improvements occur as a result of changed orientations and dispositions towards work effort (Hanushek, 1994). In another version, strongly advocated by researchers around the Consortium for Policy Research in Education (CPRE), sanctions, such as probation or reconstitution, attain their motivational power in conjunction with resources needed for capacity building in schools that fail as much for lack of will as for insufficient capacity (Fuhrman & Elmore, 2001; O'Day, Goertz, & Floden, 1995; O'Day & Smith, 1993). Thus, in the first version, clear performance goals, incentives, and sanctions make new resources unnecessary while in the second version they make new resources more effective. But in either case, the motivational power of incentives and sanctions on individuals or organizations is assumed or implied. In fact, we could probably dispense with the whole superstructure of "high stakes" that many states have built up in the last few years and return to more traditional redistributive grant making patterns (Peterson, Rabe, & Wong, 1991) if it was not for the belief in the power of incentives and sanctions for the improvement of low-performing or "failing" schools. The power of incentives and sanctions is even more crucial in systems that place the accountability burden on schools rather than districts. In such systems, incentives and sanctions must compensate for the states' limited capacity to directly regulate or administrate remote school actors.

"Naming and shaming," as the English say, threatening more severe penalties, signaling public urgency and support are major mechanisms of probation that are to impel individual educators and schools to improve. Generally speaking, while there is some research on the effect of high stakes accountability on schools generally (Firestone, Mayrowetz, & Fairman, 1998; Kelley, 1999; Kelley, Conley, & Kimball, 2000; Kelley & Protsik, 1997; Newmann, King, & Rigdon, 1997; Fuhrman & Odden, 2001), little research on the role of sanctions in low-performing schools is available (Hess, 1999; Hess, 2000; Ofsted, 1997; Reynolds, 1996; Wong, Anagnostopoulos, Rutledge, Lynn, & Dreeben, 1999; Wong, Anagnostopoulos, & Rutledge, 1998), despite the proliferation of the phenomenon. But the absence of research does not necessarily mean that "little or nothing is established," (Wilcox & Gray, 1996, p.3) or as researchers have it, that nothing "is known." While certainly holding little appeal to the education profession that is subjected to it, probation and sanctions must make intuitive sense for those that decree and design accountability systems. This article, trying to get behind these intuitions, reports on findings from a three-year study of schools on probation in the states of Maryland and Kentucky. The study investigated the effect of probation on individual performance motivation, organizational processes, and patterns of instruction. The primary purpose of this article is to provide a summary of findings in abbreviated form. (Note 1)

The Two States

High-stakes accountability has been a topic of vigorous debate and discussion among educators and educational researchers in recent years. Particularly the Texas case has found wide attention (McNeil, 2000; Sklar, Scheurich, & Johnson, 2000). The states of Maryland and Kentucky, by contrast, garnered national acclaim (Quality Counts 2001, 2001) for centering their accountability systems on tests that went beyond basic literacy and numeracy by asking students to perform complex learning, experiments, cooperative projects, complex essays, and portfolios. Although both states have by now abandoned the complex tests with which they started out, our data were collected at a time when they were still in use, though as in the case of Kentucky already contested. Thus, this study informs on schools' responses to probation in pedagogically complex accountability systems.

Naturally, there is more to an accountability system than student learning assessments. There are non-academic performance indicators (in the case of the two states mainly attendance for elementary and middle schools), there are rewards and sanctions, selection criteria for low-performing schools, exit criteria for probation, school governance requirements, planning mandates, monitoring systems, and supports for building capacity at schools. These characteristics are embedded in authority relationships between schools, districts and the state. And all of these elements are in constant flux as political coalitions shift and new plans are advanced by state policy makers (Cibulka & Lindle, 2001), making accountability systems truly moving targets of study.

At the time the study was conducted between 1997 and 2000, both states had the main features of elaborate accountability systems in common: complex student assessments, performance categories for schools, rewards and sanctions as well as school improvement planning and monitoring. But within this basic structure they differed in some respects. Compared to Maryland, the Kentucky reform was more comprehensive, more rule-bound and scripted, but also more contested and in transition (Pankratz & Petrosko, 2000). The Maryland accountability system with the MSPAP (Maryland School Performance Assessment Program) as its heart piece was more radical in its performance demands and also more consensual at the time.

Kentucky schools on probation, or in the state's language at the time "schools in decline," were identified through a straightforward formula calculated on the basis of quantitative growth expectations, and they exited the status when they met expected test score gains. Schools on all levels of absolute performance could be in decline when they did not meet their targets. Schools could attain modest, but not trivial monetary rewards for raising test scores. The state dispatched to schools "in decline" a trained change agent (called Distinguished or Highly Skilled Educator/ HSE) who provided know-how on the system's requirements and mechanisms as well as general skills in school improvement (David, Kannapel, & McDiarmid., 2000). Accountability brought to Kentucky schools more managerial autonomy from districts and a new school-internal governance structure of shared decision making and parental involvement. At the same time, the state gradually increased the prescriptiveness of the state curriculum. Towards the end of the study, the format of the central test became more traditional and sanctions lost some of their rigor, having never been fully applied anyway. The reconfiguration of the system wiped schools' slates clean again.

By contrast, Maryland left wide discretion to the state department of education in selecting "reconstitution-eligible schools" (the state's term for probation), imposing sanctions, and exiting schools from probation. The state department tended to select rock bottom performers for probation, applied final sanctions very sparingly, and set exit criteria (performance at state average) very high. State rewards and supports played a lesser role in the system. During the study period, reconstitution was a tool of the state to influence the reform disposition of two large districts that were expected to provide local resources and support to failing schools in order to avert the threat of state take-over of schools. Thus, the state exerted indirect pressure on districts to take the test seriously, but beyond that it generally provided little pedagogical guidance and capacity building. Probation evolved into a situation in which some 200 hundred schools have been languishing for years.

The Study

Findings are based on case studies of eleven schools on probation in the two states. Each case study set consists of quantitative and qualitative data: interviews, classroom observations, meeting observations, and survey questionnaires. To gain a better understanding of the behavior of a larger number of schools on probation, we also analyzed school improvement plans from 46 schools in Maryland and 32 from Kentucky as well as state test score data from Maryland. (We did not conduct such analysis for Kentucky due to the change of test formats during our study). Data collection took place between the spring of 1998 and the spring of 2000. The study investigates the role of probation in schools that serve student populations with high proportions of children from poverty and minority backgrounds. Thus, all eleven schools have high proportions of students in the Free or Reduced Lunch program. In the Maryland schools, more than 90 percent of the students are from an ethnic minority background. In the Kentucky schools, minority proportions are above the state average.

The seven Maryland and four Kentucky schools were selected according to district, school type, duration in the program, educational load, and performance history. In each state, about half of the selected schools are middle schools, half are elementary schools. In Maryland, the schools are in the two districts where almost all schools on probation are located. In Kentucky the schools reflect the state's geographic diversity. Four schools, two in each state, are probation veterans while seven schools had been identified half a year prior to data collection. We did not select schools based on their previous performance. Rather, we wanted to study the unfolding of probation, not knowing whether the schools would be successful in their improvement effort. As a result, this is not a design that allows us to evaluate the programs in the two states.

Each school was visited numerous times by at least two researchers over a two-year period from 1998 to 2000. The database for each Maryland case typically consists of a survey, a minimum of twenty-one formal, semi-structured interviews, and many more informal ones as well as six classroom observations per school. Interviewees were teachers of all subjects, administrators, instructional specialists, and other resource teachers. All principals were interviewed. We also interviewed district officials who were responsible for programs in schools on probation, as well as state officials, state monitors, and district support personnel. At least four meetings at each school were formally observed. In many cases the researchers participated in a number of additional meetings. In the Kentucky schools we interviewed slightly fewer numbers of teachers and observed fewer classrooms. Interviews were conducted with the help of standardizing protocols and transcribed and coded with the help of NUDIST. To better understand instructional patterns we analyze data from 45 (30 MD, 15 KY) classroom visits, consisting of a lesson observation and subsequent debriefing interview.

The teacher questionnaire, containing 250 items, was administered to all full time teachers at the eleven schools. Findings from the survey data stand together with qualitative data from interviews, meetings, and classroom observations. Overall response rate to the survey was 53 percent, though response rates varied by school. Across the two states and eleven schools, a total of 287 respondents returned valid questionnaires. An analysis of respondents' characteristics show that teachers with leadership roles in their schools are over-represented in the sample. However, the 200 or so interviews that were conducted with teachers from more varied backgrounds largely confirm the quantitative patterns. These interview data do not contain the bias towards teachers in leadership position.

Individual Work Motivation

Intuitions

Although I have never seen it explicitly spelled out by policy makers and designers of accountability systems why probation and the threat of sanctions would be effective motivators in educational settings, one could imagine the following intuitive scenario: When a school is publicly labeled as deficient, teachers after going through a whole range of emotions accept the urgency of improvement. This urgency is reinforced by the discomfort caused by state audits and the like. Teachers and administrators want to repair their public image, but they also take responsibility for the quality of their work. So, they take a critical look at their own work and reflect on the valid performance demands of the accountability system. They finally decide to increase effort in their own classroom and get involved in the improvement of their school. Teachers who are highly committed to their school are especially motivated. Additional support that might accompany probation is appreciated and put to good use, but fresh resources are not essential for increasing one's effort in the classroom.

Theories

In the literature, work motivation in accountability systems is often conceptualized in relationship to sense of efficacy, control, goal setting, and expectancy of rewards. These varied, though related sources of motivation are assumed to increase teachers' performance if teachers believe that the task is in their control and they have the requisite competence for its execution, if they see a connection between individual effort and expected reward and if they value the reward itself. Teachers strive for goals that are clear, specific, worthwhile, and attainable (Kelley & Protsik, 1997), and accountability systems streamline the work situation in this regard. Shamir, on the other hand, doubts the applicability of "point of action" theories of motivation, as he calls them. These models of motivation are useful in predicting discrete task behavior, but they are less powerful in explaining a "diffuse and open-ended concept of commitment" (Shamir, 1991, p. 408) that refers to a "shifting number and range of rather ill-delineated performances rather than to ironclad and numerically constant behaviors having clearly defined parameters that everyone knows" (ibid.).

In Shamir's view, expectancy and goal setting models of motivation presuppose "strong situations," i.e. situations structured by clear and specific goals, reward expectancies, and clearly identifiable relationships between increased effort, performance, and reward. Schools, however, are "weak" performance situations in which moral purpose and internalized standards are primary motivators. If he is right, then the accountability system would become motivating to the degree that it reinforces educational goals valued by teachers. But imbued in probation is not only an incentive to improve and attain rewards, either for the sake of the children or one's own professional prestige, but also an element of coercion and a threat of further penalties to which minimum compliance or exit might be the answer (Katz, 1970; Porter, Lawler, & Hackman, 1975; Vroom, 1964). While the research has not found a clear relationship between job satisfaction and work motivation, job satisfaction is related to job commitment, i.e. one's willingness to show up for work or stay on the job (Lawler, 1973; Mohrman, Mohrman, & Odden, 1996). According to Shamir, more congruence between work motivation and commitment may be expected if accountability systems tap into teachers' more deep-seated values, ideals, and performance standards.

Findings

Awareness. Probation had the attention of the majority of educators at the eleven schools, but especially in the long-term probationary schools knowledge of what the status entailed became sketchy. When the school's status was first announced in public, many teachers felt "very demoralized," "really down" or "mortified." Senior teachers who described themselves as hard-working were shocked: "I took it very personally, because of the efforts that I've made in the years that I've been here... It was almost like, I had broken an arm, and I was in a lot of pain that particular day" (A-19; eighth grade science teacher). But soon thereafter, personal distancing ensued and personal culpability was rejected: "I viewed it as a very negative cast over the school and over me, because I thought it was basically speaking about my instructional leadership. But then on reflection, I realized that it wasn't about me. ....So, once I....cleared my head of any guilt feelings, then I was able to move forward" (E-7; elementary school principal).

Mild pressure. For most interviewees, the imposition of final sanctions was inconceivable and they were not worried about their jobs. Instead, many stressed their professional worth in spite of public perceptions. On the survey, we asked teachers to rate themselves as professionals. Overwhelming majorities said they were adequately or very well prepared (90 percent), highly skilled (60 to 70 percent), and very willing to exert effort (about 70 percent), despite the fact that large numbers of teachers particularly in the Maryland schools were fairly new to the profession (46 percent five years or less) and the school (71 percent five years or less). On the other hand, our sample is biased towards activist teachers that fulfill some leadership functions at their schools (43 percent of sample). In the interviews, teachers with or without leadership positions voiced confidence and contended that they did the best they could under the difficult circumstances, such as this elementary school teacher in one of the first Maryland schools on probation: " Basically, if you think you can do it better, come in, step in, and feel free to show us how to do it any better than how we've been trying to do it.... They [the state] lay these threats on the table, 'We're gonna come take you over.'... And you just get to the point where you say....'Fine, fire me!'"

Most teachers perceived probation as mild pressure and did not worry about threats. Instead, for them probation signaled the need for support, and they were willing to endure the stigma in return for new resources: "The stigma is the minus, but the programs that come about from that is a plus. You know, it's kind of two-sided.... I think, you know, the programs that would come about because of it, you know, it outweighs the negative. I think it's good, but I think that they should get rid of the bad stigma that goes with it" (B-5). For most interviewees, probation was not occasion for self-searching. Rather it was a nuisance and stood as well for vague hopes for support.

We inquired in both survey and interviews about the accountability system in general. After all, it is this system that spells out the rules for rewards and punishments. We operationalized the theoretical models by asking how teachers perceived the importance, validity, fairness, realism, and directiveness of the system. (Note 2) Roughly speaking, the goals of the accountability system are of medium to high importance for respondents. When asked in the interviews what makes attaining high test scores important, most teachers responded that it was "a prestige thing." They didn't like "being at a school where every day in the article they say we're failures" (20-09; Kentucky middle school English), and "all the county sees is the test scores" (B-12, Maryland middle school Health). Fewer interviewees also saw the tests as a useful gauge of performance and some of them, particularly administrators, said they evaluated their success based on the scores, but for many more, the importance of the system's performance goals was connected to their concerns for diminished professional status.

Fairness. The accountability system was less connected to the quality of teaching and learning because large majorities of respondents doubted the system's fairness and validity as a gauge of good teaching. The system was seen as unfair because it did not reflect that "honestly......it's not the teaching as much as it is the children" (10-14; first grade), and "we are doing our part, ... really, the biggest part is missing for a lot of them, that's the home" (A-11; seventh grade). Many teachers rejected the burden that accountability systems singularly placed on them and called for more distributed responsibility for student achievement: "The adage that it takes a village to raise a child is true, you know, and what that accountability thing says to me is you only hold a few villagers, instead of the entire village for the accountability" (E-8, Maryland elementary school).

Validity. Only about a quarter of the survey respondents agreed that the state assessments validly reflected "good teaching." Whereas accountability systems are designed around standardized outcomes, teachers in the eleven schools were avowedly child-centered in their philosophy. Survey and interview responses were similar in this regard. "The state, I really, the test, I could care less about, to be honest with you" (B-7). For them, "just getting these kids to do their best and be able to write and to answer questions [was their] ...key priority" (B-7). In this vein, 71 percent of respondents asserted in the survey that "rather than expecting a great improvement in school performance test scores, '[they'] concentrate on individual students' growth, no matter how small." When asked to rank-order a catalog of 11 quality indicators for their work, respondents gave standardized tests ranks 7 to 9. Teachers drew their sense of success primarily from direct interaction with students, comments from parents, and teacher-made tests. Teaching "life lessons," basic skills, citizenship, and social conduct, in the eyes of interviewees essential for their clientele, were not captured by the test, but took center stage in their classrooms. Therefore "looking at the kids' background, and looking at what is written in that test and how it addresses them and the social issues that they have, they may not make that connection. So, they may not do well. But what is important to me is if my kids are learning the things that I'm teaching them, somehow they're able to connect it to the things that they're doing" (A-8; eighth grade social studies).

Thus, personalization prevailed over data-drivenness, incrementalism over the ambition of vast test-score gains, and basic skills orientation over performance-based pedagogy. Teachers' internal performance standards were not congruent with the external standards of the accountability agency. Many teachers' self-concept eschewed the image of the score maximizer in favor of the image of an educator beholden to the intellectual and social growth of individual students and committed to the needs of the local community. Likewise, rewards were derived from encounters with individual students or learning groups and from psychic satisfaction:

> I don't feel like I need to know that they think that I'm doing the best at this and they're going to reward me for this or whatever. That's just not really important to me. I like to see my students succeed and I like to think that yes, I had something to do with that. Really, that's the only reason why we're here. The other people aren't that important. It's our students that we help make some achievements. (Kentucky middle school, 40-04; sixth grade reading)

Realism and directiveness. While many teachers expected their schools to improve in the near future, optimism was much more muted when teachers rated their chances to improve according to the criteria of the accountability system. For example, on the survey 50 percent of Kentucky respondents found the system's performance goals "very unrealistic." Yet, the directiveness of the system was very high, and large majorities among survey respondents and interviewees professed to act according to the system's directions and demands. With grave doubts about the system's meaningfulness (with regard to validity, fairness, and realism), teachers said to be primarily moved out of compliance with the state and out of concern for professional prestige. Thus, in an Analysis of Variance that tested relationships between three levels of engagement in school improvement and means in the importance, validity, fairness, realism, and directiveness scales, only goal importance and directiveness show strong and significant differences in means. For work effort, a similar analysis shows moderate, but significant differences in means for directiveness, but not goal importance. Interestingly, the (reportedly) more industrious teachers were even more skeptical about the meaningfulness of the system. I conclude from these findings that the accountability system was overall a poor motivator for teachers in the eleven schools on probation, and its strongest motivators were authoritativeness and stigma producing compliance and status anxiety. Survey responses and interviews concur in this respect.

Job Commitment. While authoritativeness and stigma may have pushed people to become more involved and increase work effort, the highly motivated were not necessarily more committed to staying at their school (means on the commitment scale do not differ according to levels of engagement and effort). On many occasions, we were struck when apparent leaders in school improvement disclosed to us in the interviews that they were planning to leave. Besides better career options elsewhere, these teachers often bemoaned intolerable pressures they felt obligated to respond to at their school. But when teachers believed more strongly in the meaningfulness of the accountability system, their job commitment was significantly higher. That is, more committed teacher groups had significantly higher mean ratings on the system's fairness and realism, (but not validity). Among the various factors tested, the factor showing the strongest mean difference for job commitment was "expectation of improvement." Thus, more committed teachers were also more optimistic about their school's prospect. According to the interviews, optimism was an article of faith for some, others reasoned that as a result of probation their school would receive more attention and new resources, but almost nobody mentioned changes in their own classrooms. Improvements were mostly understood as improvement of others or the organization as a whole.

Overall, the eleven schools, particularly the seven Maryland schools, were beset with problems of teacher turn-over. About half of the respondents, again with teachers in leadership positions over-represented, were not certain about staying or were certain to leave. In fact, in many of the Maryland schools the teacher turn-over rate was about 50% from year to year. For those that planned to stay, relationships at school but also the challenge of school improvement and an optimistic outlook were drawing points, but fewer respondents named probation as a positive influence on the school. Primary reasons for Maryland teachers planning to leave were better career options elsewhere and the feeling that their school was "a sinking ship." In the four Kentucky schools, fatigue from the pressures of probation was the most important reason to leave. Given different job market conditions in the two states, it is conceivable that for the Maryland leavers exit options tempered the pressures of probation; lack of such options may have heightened these pressures for the Kentucky leavers. But it should be stressed that probation for respondents in the eleven schools resulted in mild pressure only, and given the sentiments uncovered here, an increase in pressure advocated by some policy makers frustrated with the presumed lack of effect of probation may exacerbate an already severe commitment problem in these schools.

Summary. The study found that probation under the circumstances of the eleven schools may have a weak influence on individual work motivation and an overall negative effect on teachers' job commitment. Probation in the eleven schools primarily evokes an urgency for improvements through a dynamic of compliance and concern for professional prestige. Pressures are mild, but tangible for the more obligated teachers. Although both the Maryland and Kentucky assessments are complex, they lack meaningfulness for the great majority of educators participating in the study. Teachers eschew the spirit of reward calculus, and hold against the standardized verdict of failure a philosophy of personal connection to children, psychic rewards, and incrementalism. As a result, the impetus of the system for self-examination and reevaluation of one's personal educational responsibility is weak. The mildness of pressure and the promise of new resources seem to temper potentially harmful effects of probation on job commitment, but at the same time commitment is a serious problem for the eleven schools that seems to be exacerbated by probation. With turn-over rates of 30 to 50 percent per year, and highly motivated teacher leaders leaving as likely as lesser motivated teachers, one cannot argue that with probation the "right people are leaving."

Organizational Development

Intuitions

Most accountability systems hold whole schools, rather than individuals accountable for higher performance, and it is therefore through school-wide improvement that individuals overcome the label of probation. How could probation work in this way? The label of probation throws the school in crisis, but at the same time makes people realize that "we are in this together." Intense dialog, perhaps even conflict around the discrepancy between the current situation of the school and the state's performance demands ensue. Eventually the faculty pulls together around a set of shared expectations that are the basis for a formal structure of internal accountability. Performance data bring shortcomings into focus. All parts of the school are evaluated; planning and more vigilant monitoring make the school more effective; and with determined leadership the school learns new strategies to turn itself around and change instruction.

Theories

Deliberately induced crisis and group accountability are the main motivational levers imbued in probation, but these levers often come together with programmatic and managerial mandates and supports for capacity building. This mix is to shape the organization's social interactions and improvement strategies. Organizational theorists have recognized the work group as an important source of work satisfaction, commitment, and productivity (Tannenbaum, 1970; Katz, 1970; Mohrman, Mohrman & Odden, 1996). In the field of education, studies by Rosenholtz, (1991) and Little & McLaughlin (1993) have shown that teachers increase their commitment to, and involvement in, reform when collegial relationships at school are strong, supportive, and innovative.

Very little is known about how group accountability might work in the context of schools (Hanushek, 1994; Malen, 1999). Effective schools research considers the school the most suitable strategic unit for educational improvement. Some authors believe that group accountability may be a way "to motivate teachers and administrators to enact their jobs in a manner that leads to significantly higher student achievement, *sometimes without a commensurate increase in expenditures* [emphasis added]" (Mohrman, Mohrman & Odden, 1996, p. 54). But once the attainment of rewards or the aversion of penalties is tied to the group, rewards and sanctions operate in "weaker" situations, in Shamir's terminology, as the individual reward expectations are dependent on colleagues' capacity and willingness. As a result, group-generated performance motivation must tap into teachers' more broad-based and diffuse commitment to the organization. Moreover, in the literature on high performance organizations the group is usually understood as the basis for rewards or bonuses rather than as the unit that may have to absorb sanctions and penalties. The response of work units to sanctions may flow from individual and organizational processes that are quite different from those at work in high-performance or high-involvement organizations. Responses to sanctions may be more adequately captured by a line of inquiry that places the failing organization and its crisis in the center.

Induced crisis, as a means to rouse a declining organization to focus on its essential service (Meyer & Zucker, 1989) can motivate an organization to learn (Leithwood & Seashore Louis, 1998). As probation throws schools into crisis, they unfreeze. Old routines and mental models are up for internal debate and conflict may arise. A conflict-driven scenario of organizational learning is narrated by Bennett & Ferlie (1994): "A crisis moves awkward issues up agendas...We are likely to see continuing pressure from pioneers, the formation of special groups that seek to evangelize the rest of the organization, high energy and commitment levels and a period of organizational plasticity" (p.11). In schools on probation, the initiative should move to high-performing and highly motivated, perhaps even maverick, groups of teachers and administrators. At minimum, this process entails dialog about the goals of the accountability system, a collective commitment to shared expectations, and formal structures that undergird internal accountability (Abelmann, Elmore, Even, Kenyon, & Marshall, 1999).

External threat and induced crisis, however, are not automatic triggers of learning (Levine, Rubin, & Wolohojian, 1981). According to Staw's threat-rigidity model (Staw, Lance, & Dutton, 1981), two organizational responses to threat are likely. If the group believes in the likelihood of success in meeting the new demand from the environment, increased cohesiveness, support for leadership, but also a tendency to uniformity and centralized control occurs. If the group believes in the likelihood of failure, constricted interaction gives way to leadership instability and dissension. But even organizations successfully responding to new stressful external demands tend to reinforce dominant patterns of operation, rather than learn new things, according to this model.

Findings

Our data confirm the motivational power of the group for work effort and engagement in school improvement activities. We composed from survey items several internally reliable scales that captured respondents' perceptions of their faculty: *skills of colleagues, collegiality, principal control, principal support,* and *burden sharing* (three items that specifically ask about group control and sharing the work load).[3] Across the eleven schools, respondents with reported higher levels of engagement and effort also perceived their faculties to be significantly more collegial and their principal strong, i.e. both more controlling and supportive. This contrasts with the relative irrelevance of the accountability system's meaningfulness for different levels of work motivation across the eleven schools.

A comparison between two "moving" and two "stuck" schools from Maryland (Rosenholtz, 1991) corroborates this pattern. In the two moving schools, levels of engagement and work effort as well as school-wide activity levels (identified by field work) were higher; incidentally the moving schools also managed to increase their scores on the key state assessment (MSPAP) during the study period, whereas the two stuck schools either stagnated or declined. In the two moving schools, perceptions of colleagues' skills, collegial relationships, and principal support were significantly more positive than in the stuck schools. (Principal control and burden-sharing were insignificant). By contrast, the meaningfulness of the accountability system was actually more in doubt in the two moving schools. Yet, respondents in the moving schools felt more directed by the system and attached more importance to overcoming probation and raising test scores, that is, they responded more strongly to the extrinsic motivators of the system. Thus, probation made the moving schools more moving due to higher internal organizational capacity and external pressure, i.e. stigma and authoritativeness, but not because they believed in the rightfulness of the system. But it was troubling that teachers in the moving schools were not significantly more committed to staying.

What was going on in the two moving schools? Patterns of organizational interaction and types of improvement strategies were in many respects quite similar in the two schools, one an elementary school, the other a middle schools. Both were located in the same district which had few of its schools on probation, organized a fairly efficient central Office of School Improvement, and awarded to all its "reconstitution-eligible" (RE) schools between $150,000 and $250,000 in excess of the regular budget.

Both schools were led by seasoned principals who had survived in their position, but felt nevertheless under enormous pressure. (Note 3) With teachers feeling probation only as mild pressure, organizational accountability rested on the shoulders of the principals. Their main response was to increase control and to assemble a leadership team of assistant administrators, instructional specialists, test coordinators, and school improvement resources teachers that were often hired through reconstitution-designated funds. Externally constrained by districts' and states' programmatic and managerial mandates and supports, the fate of probation, internally, was largely decided by the interplay between the principals' leadership, the skills and commitments that the specialists brought to their task, and a largely compliant, but relatively immobilized, increasingly inexperienced, and uncommitted staff.

School B was the moving middle school in the sample. According to staff comments, the RE designation made their principal into a more vigilant manager, overriding the traditional hands-off style with which administrators and staff had traditionally accommodated each other at the school. The principal abolished all faculty and team meetings and called House teams into his office once a week. During these meetings, faculty members were informed and admonished to comply with the principal's expectations and the strategies adopted by the instructional specialists. The faculty's role was to report on task completion. The principal began to visit classrooms regularly with checklists in hand. On his visits, he emphasized behavior modification strategies that could be monitored easily, such as the daily lesson plan, a fixed surface structure of the lesson, seating arrangements, bulletin board displays, the placement of the district curriculum on the teacher's desk, the page opened to the day's curriculum, etc., all of which staged the teachers' compliance with school improvement efforts. In addition, the very skillful instructional specialist had compiled a handbook of generic strategies that she believed would "crack" the complexity of the performance-based test, such as a particular surface structure for essays, a particular way of writing out a math problem, etc. During weekly campaigns, teachers were expected to practice these strategies with their students and were monitored on their use. The school raised its test scores substantially for one year, but was unable to keep up this upward trend.

Control at School B came with a smiling face. The principal was warm and paternal, but had made it clear that they, he with the rest of the school, had their backs against the wall. The instructional specialist was accommodating and always full of ideas, but the teachers knew that her proposals were what the principal wanted to get done. Teachers at School B felt controlled and supported. Many of our interviewees empathized with the principal's difficult position (it somehow reflected their own), they "understood" that accountability dictated stronger measures, and they appreciated the sense of direction that was provided for them, but at the same time many wished to escape the pettiness and pressure and work somewhere else. After the first year of probation, and despite the school's success, 70 percent of the teachers were contemplating to leave the school, and turn-over rates remained stubbornly high.

School C, the "moving elementary school" responded similarly to probation. But here the principal relied more on his team for results. One of the instructional specialists managed to develop detailed knowledge of the school's test score data and designed a daily curriculum for all lower-grade teachers. Many appreciated the support, and some were "encouraged" to teach these lessons by the principal's unannounced visits. However, although the school managed to improve as evidenced by raising test scores substantially in two consecutive years and the district alleviated severely overcrowded conditions, teacher turn over hardly abated. Out of 30 classroom teachers in 1997/98, only eight could still be found in the school at the beginning of the 2000/2001 school year; of those eight, four were kindergarten teachers. The principal had announced his retirement and the instructional specialist her leaving.

A third school, a middle school located in the same district, was also moving, but it moved astray into "pathological rigidity" as we termed it. With a legacy of discipline problems, the district installed a new principal with a proven track record of school improvement who brought with her a loyal leadership team and cleaned house. Her hallmark were a tight hall supervision policy and the same control mechanisms in use in the other two schools, minus the attention to curriculum and instruction. Debate was not tolerated at faculty meetings, and teachers' rule infractions were publicly rebuked and justified with the need for accountability, occasionally over the PA system. Test scores never improved much or declined, and year after year fifty percent or more of the faculty left. In the end the leadership team imploded and the remaining teachers threatened a walk-out.

In the two "stuck" schools, probation was an altogether less dramatic affair at the time of the study. Located in a district where support for probation schools had to be spread over half of the district's schools, Schools D and E had been designated for three years at the time of the study and probation had become habituated. If School D, a large middle school, had ever shown a more spirited response to probation, there was no trace of that during the time of our field work. MSPAP test scores had remained very low and stagnant for the entire probation period, and for teachers the signal of probation was simply submerged among the many other concerns for daily order and survival. When we first entered School D, we encountered a dispirited principal who felt he had barely made a dent in his school during his one year tenure. Frustrated by flat test scores, district inaction on the most basic building repairs, and feuding with the faculty, he was counting his days to be replaced. The next principal showed very little urgency and concern for change. He said that he would study the school the first year and then take his steps. He was liked, but he was also known to take his breaks with other teachers smoking under a tree off school grounds. His tenure ended with an acrimonious faculty meeting during which some faculty members aired their raw frustrations with his inaction.

School E, a small elementary school located in a very poor section of town, had actually been moving at some time. When the school was "named" as one of the first in the state, the school and community had organized a spirited rally in support of the school and in protest against the state. The principal, rooted in the community, had a background in staff development, but management was not her strong suit, so she opted for intensive training of her staff in performance-based pedagogy. Test scores improved remarkably and with increasing numbers of schools entering probation, she was in high demand as a speaker and trainer. But her training-based improvement model faded as the district spread staff development resources thinner with increasing numbers of probation schools, and as high teacher turn-over erased past training gains. The school was unable to fill positions, and it became common that teachers quit mid-year. While on probation, the percentage of inexperienced teachers increased to 70 percent. One year, with test scores plummeting, the school could not fill positions with permanent teachers in 3rd grade, a key testing grade for MSPAP. Also, while on probation, the percentage of special education students increased from 23 percent, already above the district average, to 27 percent. All instructional specialists had to be moved into regular instruction, and as a result the faculty according to one of the specialists did not even have sufficient basic capacity to implement the district-mandatory and very prescriptive "Open Court" literacy program even with district training. With test scores decreasing and the faculty dispirited and worn-out, the district decided to replace the principal who hitherto had provided a modicum of stability, with a new, but inexperienced principal. The school declined even further, and after six months into the tenure of the new principal, the state announced that the school would be taken over by the state the following year. That year, prior to the actual take-over, satisfactory scores on the MSPAP plummeted to zero or near zero.

Summary from the Maryland Cases

Rather than staging crisis and opening channels of inquiry into solutions with broad faculty participation, administrators are *in* crisis and as conduits of accountability tend to mute the voices of outspoken critics who might question the undisputed reality and legitimacy of the accountability system, but whose ardor might also expose the school to honest self-evaluation. Accountability is accepted as a fact; the value and realism of performance goals is not publicly deliberated in most schools. The teachers are willing to rally around their leader as long as they sense tangible progress. Teachers resist crude managerial control, but accept increased control in those schools where it is laced with traditional paternalism and concrete assistance. Teacher learning takes place as skill (re)training primarily.

High principal turnover or low-impact principals doom a school's probationary period. Our "success" cases have higher (perceived) capacity. They are more unified and supportive, their faculty is perceived as more skillful, capacity building is seen as more effective, and the district is more forthcoming with new resources and interventions. But also, probation makes teachers compliant, and traditional prerogatives of teachers' classroom autonomy are overcome through administrative power attached to specialists' instructional support. In the more successful cases, increased rigidity is associated with more effectiveness of the organization. Discipline tightens; more attention is paid to the state assessments; classroom teachers are on guard. Career teachers and instructional specialists are roused into action and rally around the principal. A curriculum is being followed. Increased participation in staff development workshops may have increased the competence of (especially novice) teachers. Some of the seven schools post modest improvements in this way. But increased organizational rigidity exerts a price. Teachers are dissatisfied; some resent the pressure and standardization, and many contemplate leaving. These are not circumstances under which internal accountability flourishes. Improvement strategies chosen by the schools correspond to the patterns of leadership. Schools rely on external programs, sweeping standardization, easily surveillable behavior, surface compliance reviews, and test preparation schemes (see also Mintrop & MacLellan, 2002, for results of the content analysis of school improvement plans).

Evidence from the Kentucky Cases - Rethinking the Pattern

On the individual level, notwithstanding different weights, basic patterns were similar in schools from both states. But on the organizational level, we did not observe the same kind of organizational rigidity pattern in the four Kentucky schools that were so prevalent in the Maryland schools. To begin with, probation in the Kentucky schools was an altogether less stirring affair. (This is 1998 when the first wave of accountability demands is spent and the system is in the throes of political contestation.) Teachers and administrators stressed continuity of their school's efforts to improve regardless of the school's status. The Kentucky respondents attached less importance to higher test scores and less meaning to the accountability system. Performance problems tended to be externalized. One school considered itself the "district dumping ground," another the district "special ed magnet." Although public stigma hurt and instilled in most teachers a desire to shed the "in decline" label, they reported to a lesser degree than Maryland teachers having exerted more effort as a result of probation.

Although, compared to Maryland responses, Kentucky respondents were less optimistic about their school's prospects of improvement and less certain about their efficacy with their students despite higher levels of work experience, they gave their schools higher marks on capacity. Faculties were seen as more skillful and collegial and principals as more supportive and less controlling. Principals themselves did not feel threatened in their jobs based on test scores. Three of the four principals owed their long tenure to districts that the state accountability system largely by-passed, and district interventions were not as prominent. Hence the urgency that fueled control strategies in the Maryland schools was largely absent. Teachers felt challenged to do a better job at aligning their curriculum with the increasingly prescriptive state core curriculum and pay more attention to test-specific features, such as writing prompts. Because the faculties were fairly stable, there was more evidence of training effects in the interviews. The Distinguished/ Highly Skilled Educators provided assistance in assessment-specific features and were seen as helpful in keeping their schools focused on what schools could internally control (Mintrop, MacLellan, & Quintero, 2001). But they did not direct the schools' improvement strategies. Rather their formal authority position was absorbed into the traditional hierarchy of the schools. More managerial autonomy of the school fostered entrepreneurialism in attracting new grants and projects. "You name it, we've tried it," as one principal termed it, was the visible badge of the schools' commitment to improvement. Several HSE's bemoaned that this approach left (low) expectations and classroom routines in these schools largely unexamined.

In summary, then, the detected rigidity effects of probation in the Maryland case may not be a general pattern of response to probation, but be related to a specific constellation of factors: More district control, threatened principals, and ordinary teachers with low skills, low commitment, and modest work motivation all working within a state accountability system that steers local districts with pedagogically complex outcome demands without providing the tools to reach them. Thus, one might say that the Maryland schools are a case of high administrative pressure meeting low capacity. In the four Kentucky schools, we observe a more traditional pattern of school improvement through alignment and the acceleration of add-ons. The eleven schools have in common, however, the absence of dialog about teachers' responsibility and the school's expectations, and a conversation about a meaningful response crafted in the tension between the school's and the accountability system's shortcomings.

Instructional Changes

Intuitions

Given the ambitious performance-based character of the accountability systems studied here, schools, in order to master probation successfully, not only need to compel students to work harder, but also learn differently. Higher work intensity, tighter lesson plans, but also higher order thinking and teamwork are paramount. When teachers have the will to change and faculties have begun to evaluate the shortcomings of their school, raise their own expectations to the high demands of the system, and agree on formal procedures of internal accountability, the conditions are ripe for a restructuring of teaching content and methods.

Theories

The literature on curriculum policy and instructional change shows that what teachers learn from policy depends on a host of factors: their extant practices, their understanding and interpretation of the policy, their own experiences, dispositions and skills, and the support they receive in efforts to change their practices (Cohen, McLaughlin, & Talbert, 1993; Darling-Hammond, 1997; Grant, 1998) found that teachers responded quite differently to the same reform, even when exposed to the same interventions. Spillane & Jennings (1997) show that when districts employed alignment strategies to change instructional practice at the level of classroom discourse, they often achieved superficial task modification, but did not reach more deeply ingrained task and discourse structures. Two responses are observed in the literature. Teachers often trivialize complex tasks to simpler task demands (Cohen, 1990); Spillane & Jennings, 1997; see also Spillane & Zeuli, 1999; Cohen & Ball, 2001) and they doubt the relevance of ambitious performance standards and institutional demands when incongruent with the perceived needs of their students (Darling-Hammond & Wise, 1985). In this case, institutional demands and school reality come in conflict with each other unless high standards are examined in light of real student work (McDonald, 1996). If teachers learn ambitious pedagogy through "revisiting and reinventing" (Cohen & Ball, 1999), then probation cannot succeed without accountability being connected to personal educational meanings and processes of organizational dialog and learning that facilitate exploring these meanings.

Findings

It is apparent from the previous sections on individual learning and organizational development that probation in the context of the eleven schools provided unfavorable conditions for learning new and ambitious performance-based pedagogy. For many teachers, the state assessments did not provide meaningful tools for the self-evaluation of their teaching. Teachers were not data-driven. Rather data from the state assessments tended to be discounted in their value and validity. On the organizational level, probation fostered rigidity and compliance with external obligations, as in the case of the Maryland schools, to the detriment of organizational learning and internal dialog. Moreover, whereas the accountability system calls for an upgrading of teaching quality, the investigated schools on probation struggled with high teacher turnover, low job commitment, an increasing number of uncertified and inexperienced teachers, and in some cases highly unsupportive districts.

We saw that large numbers of teachers in the 11 schools viewed themselves as highly competent professionals whose skills and knowledge measured up to the demands of the states' performance-based assessments. But in reality, 70 to 80 percent of the observed lessons in Maryland did not show evidence of elaborate level teaching at all; that is, the frequency of higher-order thinking, problem solving, and complex dialog among the counted snapshots (total number 150) was very low. (Due to their limited number, we did not quantify the observed lessons in Kentucky.)Only one third of all observed Maryland lessons were deemed highly coherent, i.e. beginning, middle, and end hung together; the majority lacked conceptual depth. On the positive side, in the overwhelming majority of lessons, teachers used a variety of materials, activities, and forms of interaction; in quite a few lessons variety was a very prevalent feature. Contrary to some other assessment systems that emphasize minimum competency tested in a multiple-choice format (Darling-Hammond, 1991; Noble & Smith, 1994), evidence of practicing simple test taking skills (i.e. "drill and kill") was fairly low. In all likelihood, the complexity of the state assessments in the two states did not lend itself to such an approach.

A selection of seven teachers illustrates patterns observed across the 30 classrooms. We observed the classroom of a senior middle school teacher who had the reputation as an innovator. In the observed lesson, she had students measure the relationship between diameter and circumference of various circular objects. By following the lesson "script" from the newly adopted mathematics textbook, she believed that her instruction was aligned with MSPAP because "they match the skills with the national standards. So it is really close because I know the MSPAP is taken from the state standards that they get from the national standards" (G-16).

The problem was that, contrary to the intentions of the book, she herself introduced Pi without the students having had a chance to discover the relationship themselves. She said, "I'm a math teacher. I'm used to, you know, this is this and this is this..." To her, the ability to work successfully in groups was the key to MSPAP proficiency, construction of concepts was the lesser of her concerns. But despite her willingness to change, accountability for her was "treating you more like a child... So, I see it as the work doubling....And we do it, I mean, you know, they say "this" and we do it. [But] the morale doesn't work very well. (G-16)

A young and effective mathematics teacher in another middle school coped with accountability differently. Because of the variety of testing situations she had to prepare her students for, she parsed her lessons in a regular pattern of basic skills, performance-based, and regular lessons: "If it's just a regular lesson, no MSPAP, no Functional per se, then I rely on the book" (B-17). She was obviously a skilled classroom manager and adept at teaching. When asked how she envisioned closing the gap between her students' capacities and the state's expectations, she responded: "Well, one day at a time, basically. I knowthat the gap cannot be tightened within a, you know, short period of time... and in time, if instruction is, if you're doing what you're supposed to do, then the scores will come up" (B-17)

We visited the classroom of a very respected science teacher with the reputation of a disciplinarian. He taught a very directive and repetitive lesson. Changing his teaching in response to the accountability system was out of the question for him. He taught the content he believed needed teaching, with the materials of his choice (very old textbooks), in the manner that he saw fit. An alternation of very directed reading with experiential lessons worked best for him. Similarly, his colleague, a middle-aged woman, highly respected by students, taught a very traditional lesson that kept students working hard. She introduced herself: "I don't know if you want this on tape but I have a sticker on my car that says, 'Stop MSPAP, teach basics.'" She subverted external pressures and was outspoken about her conviction that she had better sense than the various distant agencies and actors that tried to tell her what worked with her students:

> No matter what happens, I don't change, and they [the students] depend on it. That's important to them. It's also a part of classroom climate. My expectations don't change for them. They know what to expect. (A-22)

An elementary school teacher, in contrast, was delighted with the instructional materials mandated by the school. She loved the scripted nature of the *Open Court* reading. She was confident that if students could read they would be successful at taking the MSPAP test when the time came, or any other test for that matter.

Mr. C. faced the problems that many beginning teachers face. We observed a frustrating lesson during which he attempted to teach the difference between "action and state of being verbs." He received little instructional support or guidance. Student discipline was not, in his view, an administrative priority at his school and that combined with parental non-involvement made classroom management very difficult for him. He faulted his inability to raise his instruction to grade level on his students' lack of knowledge. Mr. C. assured us that "I can be rather creative when I'm in the right environment." When asked whether the reconstitution eligibility status of the school or the MSPAP influenced his teaching, he replied, "Is MSPAP driving what we're doing in any way? No. No. What's driving what we're doing is survival" (G-21). Another beginning teacher, Mr. S., was in the same situation. Nonetheless, he made an effort to follow the adopted curriculum as closely as possible, but was often thrown for a loop, as in the observed lesson, because the curriculum did not match up with his students' below-grade-level skills. He did not think that students could come up to the expected performance level because his third grade students were "already so far behind" (F-20). Mr. S. was, in his words, "worn out" and "run down" by trying to reconcile the reality that "a lot of things that the kids come into school with are things that are way beyond my control" and being held accountable for student achievement. He intended to leave teaching at the end of the school year.

Most teachers we visited considered it unlikely that their students would reach the lofty goals the accountability had set out for them. But teachers were willing to try concentrating on incremental learning steps. In negotiating the gap between external performance demands and the perceived abilities of their students, teachers foremost gauged their lessons to students: "I teach to the needs of the students. That's what I feel I should do because if I try to teach up here and they're not up there, I'm wasting my time and my energy because they will never meet with success" (F-17; second grade). Teachers felt justified teaching lessons in a basic skills format that traditionally "worked" for "their kids." In the view of many, MSPAP activities distinguished themselves mainly as writing activities, group work, and the use of particular analytic vocabulary. For fewer teachers, reflection on one's own thought process was also associated with MSPAP. This pattern holds across all observed teachers. Often the conceptual depth of knowledge construction that is a core element of the new pedagogy was simplified into a set of activity formats. Judging from the debriefing interviews that accompanied lesson observations, teachers were, for the most part, not aware of this task trivialization. This was not surprising, considering that the test itself was shrouded in mystery and teachers only reluctantly discussed items that they had seen for fear of doing something inappropriate. Thus, while teachers on one hand did not reach the levels of pedagogical complexity that the state assessments envisioned, test practice was not trivialized to the level of learning how to "fill in the bubbles" either. The more performance-based assessments, in place in Maryland and Kentucky at the time of the study, may have discouraged this.

Although teachers strongly expressed the notion that their lessons were first and foremost adapted to their students' ability and achievement levels, tests, new instructional programs, new curricula, and new textbooks reached deeply into many teachers' classrooms. But external pressures and directions were multiple and often contradictory. In the survey, more than half of the teachers felt clearly directed by the accountability system. Observations show this clarity much more laden with conflict.

For all their resentment, many teachers, almost in passing, expressed habitual compliance with administrative mandates intended to align instruction with MSPAP. Although they saw the accountability system as unhelpful and stacked against them, they did not reject it and did not outright condemn it. They truly served two masters. They wanted to concurrently accept the institutional weight of the state and be sensitive to the needs of their students, but the two pulled from opposite ends. Some teachers learned from this tension, but more frequently tension was diffused by discarding the state's directives by virtue of their unreasonableness, or by discarding students as uneducable. But the great majority adopted officially sanctioned programs as a defensive retreat that relieved them of dissonance and delegated the decisions and responsibilities to a higher level. In all the schools, a main feature of instructional reform was the monitoring of surveillable behavior. A few teachers considered this "good pressure" (F-18), but many experienced teachers thought it failed to attack the real problems faced by the school, like this senior teacher:

> We are now held to certain standards and expectations such as we have expectations for the students, the principal has some and administration has expectations for us and she checks to make sure that these things are being done....In the past we've more or less been left to do to our own task and like I used to be guilty of not writing lesson plans. I'd come in, I knew what I wanted to teach....Really, the only thing I do differently now is write out.....But other than that, there's nothing I didn't do that I've changed. I still have my objectives, my outcomes, my warm-ups....That was all there before. (A-24)

Instructional Change in the Four Kentucky Schools

Descriptions of Kentucky classrooms are based on a smaller base of observations. Therefore Kentucky patterns are more anecdotal. None of the Kentucky teachers we visited in their classrooms reported instructional sea changes as a result of the state assessments. Most teachers described their instructional changes as "add[ing] skills here and there" (10-10, second grade math). Patterns of instruction in the small number of classrooms we visited in the Kentucky schools were very similar to the patterns encountered in the Maryland schools. Most lessons were taught on a basic level, some lessons were marginal.

Changes mentioned by teachers bifurcated. There were those that had to do with alignment. Overall Kentucky interviewees seemed relatively well informed about unique features of the system, most notably portfolios and writing prompts. Conceivably, with less staff fluctuation, professional development and assessment-related training may have left their mark to a much larger degree than in the Maryland schools. Highly Skilled Educators cautiously focused schools on key tasks of the reform, such as planning, portfolios, writing, and curriculum alignment. Most teachers said they tried to cover the core curriculum as best they could:

> Well, each school does an aligned curriculum, so that's what I'm supposed to teach. That's my aspect of it....What we try to do is make sure that we have given them a thorough review for the test. We try to get as much through as possible. With the test, they give you roughly what percentages, like 10 percent is going to be weather and stuff like that, so you say, "Well, okay, it's going to be 10 percent weather, so I can give them worksheets on weather and give them a project on weather." Things like that....(40-15, sixth grade science)

But on the other hand, Kentucky schools had wide discretion in selecting programs and strategies. For some teachers, instructional goals were to raise achievement on state assessments while many others revealed that the need to motivate students was a more persuasive influence on their decision making about instructional materials. Pressures from the school's accountability status did not seem to foreclose their own approaches and curriculum alignment left flexibility. Overall, being "in decline" did not exert a strong press towards broader instructional changes. One teachers weighted the effect of the "in decline" status this way:

> Yes and no. Yes because there are certain things that I must do to... They have requirements now for us, like we have to do so many questions with the kids, like open response questions and things like that, so yes, it affects me in that way. The curriculum is not changed any. The curriculum is the same. If they were not here, if I'm doing my job, then I'd still be doing the same curriculum. Channel 11 was in here and they asked the same question. They said, "What are you doing...blah blah blah?" And I said, "Well, the academics are still the same. It's just our accountability that's up for grabs right now." We have to show that we're doing these things by performing, jumping through certain hoops. But no, I don't think I'm a better teacher because I'm in decline trying harder.....I don't think I am doing anything different. I do post stuff, though, because I'm supposed to. .....I didn't have all the nice little things that they give us, so now I can label everything I do.... We have to succeed on a test that the students take. .....The students will take the test, but their grades are going to have to be high enough that we succeed. (40-16)

Neither in the Maryland nor in the Kentucky classrooms did we encounter much of the mind-numbing test drill and practice that have been reported from accountability systems in which traditional basic skills tests have become high-stakes.

The Limits of Sanctions

Incentives and sanctions are the linchpin of a new generation of high-stakes accountability policies. I have explored how schools in pedagogically complex accountability systems responded to the signal of probation. On the positive side, almost all of the eleven schools were modestly energized by the label, at least at some point or from time to time. Teachers in all schools reported that they increased work effort and engagement in school improvement as a result of pressure and direction. Management in some schools tightened up, educators paid closer attention to the state assessments, support from instructional specialists intensified, and the adoption of new programs, strategies, and projects accelerated. In this way, a number of schools were able to remedy some inefficiencies and provide more structure to teachers than previously had been there.

A look at MSPAP test scores across the seven schools from Maryland during the post-probationary period suggests as much. While two schools have made notable strides on the MSPAP test in the areas of math and reading since becoming "reconstitution-eligible," for the majority MSPAP performance in these key areas has either seen a very modest increase or decline, but mostly stagnancy. Additionally, schools have been plagued with year-to-year score fluctuations. But all seven schools arrested decline in the first years after identification. In all these respects, our sample resembles the overall patterns identified for all Maryland schools on probation in the 1996 and 1998 cohorts. We concluded from this analysis that probation may foremost be a tool to arrest decline in persistently low-performing schools, but it may not produce large gains. By remedying gross inefficiencies, many schools are able to "harvest the low-hanging fruit," as one of my colleagues calls this stage, but they make few further inroads into the territory of instruction.

Had the assessments been less complex and more basic skills oriented, probation may have "worked" better, that is, the pressure of the stigma combined with various control strategies and program standardization I described earlier could have produced an intensification of instruction based on already existing competencies and instructional formats. But pressure is a double-edged sword. It may challenge people to increase work effort, but also make them want to leave if they do not value the pressure as serving a worthy purpose. Increasing pressure would have exacerbated the already immense problem of job commitment in the studied schools. Moreover, teachers knew that they did not face too much competition for work places that many outsiders viewed as unattractive. Hence, as to sanctions, most teachers called the states' bluff. Principals, however, standing in for the accountability of the organization, felt differently.

Probation was not working well as a tool for instructional reform. To begin with, majorities of teachers did not find the standards of the accountability systems meaningful for their work as educators. Being "unfairly" branded as "deficient" by a system that was seen as insensitive to the needs of their disadvantaged students, they in turn based their own sense of worth on personal relationships in their close-up environments. Rather than accepting criteria and judgments of the system, they felt singled out as the ones who had to carry the "blame" for student learning and in turn externalized the causes for underperformance. Probation did instill in schools the notion that "something" had to be done, but in none of the schools did probation trigger elements of internal accountability, if this is to mean a process through which a faculty formulates its own expectations in light of student needs and high demands of the system, agrees upon formal structures that hold them to account, and focuses improvement on identified key instructional deficiencies.

This kind of internalization process was neglected in the Kentucky schools, but forestalled in the Maryland schools. The rigidity pattern encountered in the latter is an example of what happens when high performance demands and top-down pressure meet low capacity schools. The result was a proliferation of control strategies that had the potential to turn classrooms into the opposite of what performance-based pedagogy intended. Being "in the fish bowl" most teachers tightened up traditional lesson structures. Coverage and task completion reigned supreme and more group work and writing assignments were added. Looked at through the perspective of the seven focal schools, the Maryland case illustrates the limits of steering educational reform through incentives and sanctions placed on outcomes without an instructional technology that facilitated the alignment of demanding outcomes with curricula and materials and provided a bridge to student needs. The state also left capacity building largely in the hands of local districts, with the result that external demands and pressures fell upon wholly unprepared schools that reacted with rigidity, rather than learning. In the Kentucky cases these responses were avoided. Here, by contrast, higher capacity schools responded to a system that was less ambitious pedagogically, more prescriptive as to alignment, and more supportive through the Highly Skilled Educator feature.

But there are success cases among the Maryland schools on probation. And one of the moving schools in our sample may give us an idea of what went into their improvement: an experienced principal, exceptional instructional specialists with data analysis, curriculum development, and coaching abilities, and additional resources provided by the district. And yet, prospects for the school are dimmed. Leaders exit, and without a process that involves all faculty members as responsible and committed actors, rather than mere implementers, the school may find itself at a loss again.

The accountability systems in both states operate on the assumption of organizational stability. Only this assumption makes it legitimate to publicly expose putative deficiencies of whole schools based on year-to-year comparisons of schoolwide test scores. The reality of the 11 schools on probation selected for the study is, however, quite different. These schools were, on the whole, hard and challenging work environments that educated large numbers of students considered at risk. Particularly the 7 Maryland schools faced instability due high teacher and administrator turn-over, and increasing proportions of inexperienced teachers which first and foremost raises the specter of student discipline problems and social instability. But the Kentucky schools as well were beset with high student mobility rates and changing student in-take or attendance zones, changes that were outside of their control. Under these conditions, continuous improvement was an impossibility due to the schools' lack of organizational continuity. Thus, many of the schools, particularly in Maryland, needed baseline stabilization first before they could embark on ambitious instructional reforms. The responsible actors for this kind of stabilization are for the most part districts and states. Teachers themselves give us an idea of what is urgently needed. When asked to select among 10 priorities for school improvement, large majorities chose student discipline, teacher motivation, and teacher turn-over while not even ten percent believed that a new pedagogy should be first on the agenda.

Note: I would like to thank Daria Buese, Kim Curtis, Masako Nishio, Lea Plut-Pregelj, and Margaret Quintero for their assistance, as well an anonymous reviewer for very insightful comments. This research was supported by a grant from the Office of Educational Research and Improvement, U.S. Department of Education (No. R308F70035). The views expressed in this article are mine.

Notes

1. See the full report at www.gseis.ucla./faculty/mintrop, "Schools on Probation."
2. These were direct questions in the interviews and scales derived from questionnaire items. (See Technical Report for details at www.gseis.ucla.edu/faculty/mintrop/).
3. Since many reconstitution-eligible (RE) schools in Maryland (including the seven selected schools) improved only marginally or not at all after identification, punitive transfers of principals were frequent. In four of the seven schools in our selection, the RE designation was accompanied with an immediate change of the principal. Two of the four new principals did not survive their first year after RE designation, one was transferred after his second year. One school had a new principal every year for the three years of data collection. In three schools, the long-term principals survived the RE designation, but they felt highly uncertain in their tenure. One of them subsequently lost her job and chose early retirement, leaving only two principals who survived RE designation in their assignments. One of those two retained his job against the explicit wish of the state department to remove him and one retired two years after his school's probation designation. By comparison, across the four Kentucky schools, the situation was more stable. Three of the four schools were headed by principals with long tenure in their schools. One school, by contrast, had a new principal every year in the last six years, though this was not attributed to the school's performance status, but to district problems.

References

Abelmann, C., Elmore, R., Even, J., Kenyon, S., & Marshall, J. (1999). *When Accountability knocks, will anyone answer?* (CPRE Research Report Series RR-42). Philadelphia: Consortium for Policy Research in Education - CPRE.

Bennett, C., & Ferlie, E. (1994). *Managing crisis and change in health care: The organizational response to HIV/AIDS.* Philadelphia: Open University Press.

Boser, U. (2001, January 11, 2001). Pressure without support. *Education Week, 20,* 68-71.

Chicago Public Schools (1997) *School Probation Information Packet.* Chicago: Office of Accountability.

Cibulka, J., & Lindle, J. (2001). *The politics of accountability for school improvement in Kentucky and Maryland.*

Technical Report, Volume IV. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Cohen, D., & Ball, D. (1999). *Instruction, capacity and improvement*. Philadelphia: Consortium for Policy Research in Education (CPRE Research Report).

Cohen, D., & Ball, D. (2001). Making change: Instruction and its improvement. *Phi Delta Kappan, 83*(1), 73 - 77.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation & Policy Analysis, 12*, 311-329.

Cohen, D. K., McLaughlin, M. W., & Talbert, J. E. (1993). *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.

Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. *Phi Delta Kappan, 73*, 220-225.

Darling-Hammond, L. (1997). School reform at the crossroads: Confronting the central issues of teaching. *Educational Policy, 11*(2), 151-166.

Darling-Hammond, L., & Wise, A. E. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal, 85*(3), 315-336.

David, J., Kannapel, P., & McDiarmid., C. (2000). *The influence of distinguished educators on school improvement: A study of Kentucky's school intervention program*. Lexington, KY: Partnership for Kentucky Schools.

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95-113.

Fuhrman, S., & Elmore, R. (2001). Holding Schools Accountable: Is It Working? *Phi Delta Kappan, 83*(1), 67-72.

Fuhrman, S., & Odden, A., R. (2001). Introduction: A Kappan Special Section on School Reform. *Phi Delta Kappan, 83*(1), 59-61.

Grant, S. (1998). *Reforming reading, writing, and mathematics: Teachers' reponses and the prospects for systemic reform*. Mawah, NJ: Lawrence Erlbaum.

Hanushek, E. A. (1994). *Making schools work: Improving performance and controlling costs*. Washington, D.C.: Brookings Institution.

Hess, G. A., Jr. (1999a). Expectations, opportunity, capacity, and will: The four essential components of Chicago school reform. *Educational Policy, 13*(4), 494-517.

Hess, G. A., Jr. (2000). *Changes in student achievement in Illinois and Chicago, 1990-2000* (An occasional paper released by the Brown Center on Education Policy). Washington, D.C.: The Brookings Institution.

Katz, D. (1970). The motivational basis of organizational behaviour. In V. Vroom & E. Deci (Eds.), *Management and motivation* (pp. 265-292). Hammondsworth, GB: Penguin.

Kelley, C. (1999). The motivational impact of school-based performance awards. *Journal of Personnel Evaluation in Education, 12*(4), 309-26.

Kelley, C., Conley, S., & Kimball, S. (2000). Payment for results: Effects of the Kentucky and Maryland group-based performance award programs. *Peabody Journal of Education, 75*(4), 159-199.

Kelley, C., & Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. *Educational Administration Quarterly, 33*, 474-505.

Lawler, E. (1973). *Motivation in work organizations*. San Francisco: Jossey-Bass.

Leithwood, K., & Seashore Louis, K. (Eds.). (1998). *Organizational learning in schools*. Lisse, NL: Swets & Zeitlinger.

Levine, C. H., Rubin, I. S., & Wolohojian, G. G. (1981). *The politics of retrenchment: How local governments manage fiscal stress*. Beverly Hills: Sage.

Little, J. W., & McLaughlin, M. W. (1993). *Teachers' work: individuals, colleagues, and contexts*. New York: Teachers College Press.

Malen, B. (1999). On rewards, punishments, and possibilities: Teacher compensation as an instrument for education reform. *Journal of Personnel Evaluation in Education, 12*(4), 387 - 94.

McDonald, J. (1996). *Redesigning school*. San Francisco: Jossey-Bass.

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

Meyer, M. W., & Zucker, L. G. (1989). *Permanently failing organizations*. Newbury Park: Sage.

Mintrop, H., & MacLellan, A. (2002). School improvement plans in elementary and middle schools on probation. *Elementary School Journal, 102*(4), 275 -300.

Mintrop, H., MacLellan, A., & Quintero, M. F. (2001). School improvement plans in schools on probation: A comparative content analysis across three accountability systems. *Educational Administration Quarterly, 37*(2), 197 - 218.

Mohrman, A. M., Jr. , Mohrman, S. A., & Odden, A. R. (1996). Aligning teacher compensation with systemic school reform: Skill -based pay and group-based performance rewards. *Educational Evaluation and Policy Analysis, 18* (1), 51-71.

Newmann, F. M., King, M. B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review, 67*(1), 41-74.

Noble, A. J., & Smith, M. L. (1994). Old and New Beliefs about Measurement-Driven Reform: "Build It and They Will Come". *Educational Policy, 8*(2), 111-136.

O'Day, J. A., Goertz, M. E., & Floden, R. E. (1995). *Building Capacity for Education Reform* . New Brunswick: Consortium for Policy Research in Education, U.S. Department of Education, Office of Educational Research and Improvement, Educational Resources Information Center.

O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. H. Fuhrman (Ed.), *Designing coherent education policy: Improving the system* (pp. 250-312). San Francisco: Jossey-Bass.

Ofsted. (1997). *From failure to success: How special measures are helping schools improve* . London: Office of Standards in Education.

Pankratz, R. S., & Petrosko, J. M. (Eds.). (2000). *All children can learn: Lessons from the Kentucky reform experience*. San Francisco: Jossey-Bass.

Peterson, P., Rabe, B., & Wong, K. (1991). The maturation of redistributive programs. In A. Odden (Ed.), *Education policy implementation* (pp. 65-80). Albany, NY: Suny Press.

Porter, L., Lawler, E., & Hackman, R. (1975). *Behavior in organizations*. New York: McGraw-Hill.

Quality Counts 2001. (2001). Quality Counts 2001: A Better Balance. *Education Week, 20*(17), 8-193.

Reynolds, D. (1996). Turning round ineffective schools. In D. Gray, D. Reynolds, C. Fitz-Gibbon, & D. Jesson (Eds.), *Merging Traditions: The Future of research on school effectiveness and school improvement* . London: Cassell.

Rosenholtz, S. (1991). *Teachers' work place: The social organization of schools*. New York: Teachers College Press.

Shamir, B. (1991). Meaning, self, and motivation in organizations. *Organization Studies, 12*(3), 405-424.

Sklar, L., Scheurich, J., & Johnson, F. (2000). *Equity-driven, achievement-focused school districts: A report on systemic school success in four Texas school districts serving diverse student populations* . Austin, TX: Charles A. Dana Center.

Spillane, J. P., & Jennings, N. E. (1997). Aligned instructional policy and ambitious pedagogy. *Teachers College Record, 98*, 449-481.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reform. *Educational Evaluation and Policy Analysis, 21*, 1-27.

Staw, B. M., Lance, E., & Dutton, J. (1981). Threat-rigidity effects in organizational behavior: A multilevel analysis. *Administrative Science Quarterly, 26*, 501-524.

Tannenbaum, A. (1970). The group in organizations. In V. Vroom & E. Deci (Eds.), *Management and motivation* (pp. 218-225). Harmondsworth, GB: Penguin.

Vroom, V. H. (1964). *Work and motivation*. New York: John Wiley & Sons, Inc.

Wilcox, B., & Gray, J. (1996). *Inspecting Schools: Holding schools to account and helping schools to improve.* Buckingham: Open University Press.

Wong, K., Anagnostopoulos, D., Rutledge, S., Lynn, L., & Dreeben, R. (1999). *Implementation of an educational accountability agenda: Integrated governance in the Chicago public schools enters its fourth year* . Chicago: University of Chicago. Department of Education.

Wong, K. K., Anagnostopoulos, D., & Rutledge, S. (1998). *Accomodation and conflict: The implementation of Chicago's probation and reconstitution policies* (Paper presented at the 1998 Annual Meeting of the Association for Public Policy Analysis and Management). New York: Association for Public Policy Analysis and Management.

About the Author

Heinrich Mintrop
Graduate School of Education & Information Studies
3335 Moore Hall
Box 951521
University of California, Los Angeles
Los Angeles, CA. 90095

Ph: (310) 794-7477
Fax (310) 206-6293
E-mail: mintrop@ucla.edu

Heinrich Mintrop was a teacher in both the United States and Germany before he entered into his academic career. Currently he is an assistant professor at the University of California, Los Angeles. He explores how education policies form institutional structures that in turn shape teaching and learning in schools. He is particularly interested in the tension between student achievement and citizenship, accountability and democratization in educational systems across the world. He recently published "Changing Core Beliefs and Practices Through Systemic Reform: The Case of Germany After the Fall of Socialism" in *Educational Evaluation and Policy Analysis* (1999) and co-authored with Bruno Losito, CEDE Italy, "The Teaching of Civic Education," a chapter in the IEA Report on *Civic Education in 28 Countries* (2001). In recent years, Dr. Mintrop has turned to the issue of school accountability with emphasis on policies that support school improvement in low-performing schools. This work has so far produced a number of articles and book chapters. A book at Teachers College Press on *Low-Performing Schools Facing High Stakes* is forthcoming.

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov

California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 3397 times since January 29, 2003

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Education Policy Analysis Archives
Volume 11 Number 4 **January 29, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the **EDUCATION POLICY ANALYSIS ARCHIVES**. Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Varieties of Knowledge and Skill-Based Pay Design: A Comparison of Seven New Pay Systems for K-12 Teachers

Anthony Milanowski
University of Wisconsin-Madison

Citation: Milanowski, A. (2003, January 29). The varieties of knowledge and skill-based pay design: A comparison of seven new pay systems for K-12 teachers, *Education Policy Analysis Archives*, *11*(4). Retrieved [date] from http://epaa.asu.edu/epaa/v11n4/.

Abstract

This article describes the design of knowledge and skill-based pay systems for K-12 teachers in six U.S. school districts and one charter school. Based on a theory of action that relates knowledge and skill- based pay systems to improvements in instruction, and the expectancy theory of motivation, seven dimensions for comparison are identified and the systems are compared based on these dimensions. While there were a variety of reasons for designing new pay systems, similarities included that teachers were involved in the design processes, and that the knowledge and skills rewarded are more closely related to instruction than in the traditional salary schedule (though none of the systems placed heavy emphasis on content-specific pedagogy). Most systems made use of existing standards or definitions of good teaching, such as the Framework for Teaching (Danielson, 1996). While most of the systems involved performance-based assessments of teacher skills, in no case were seniority and graduate degrees eliminated as a basis for pay progression. Few of the programs had developed a coordinated professional development program specifically linked to the knowledge and skills rewarded by the new pay system. Implications for policy makers and system designers are drawn.

Introduction

A number of lines of research (e.g. National Commission on Teaching and America's Future, 1996; Slavin and Fashola, 1998; Wright, Horn, and Sanders, 1997; Bembry et al, 1998; Ferguson and Ladd, 1996) have identified teacher instructional capacity as a key variable in the success of educational reforms in improving student achievement. For the past two years, the CPRE Teacher Compensation Project has been studying a new form of teacher compensation that may have the potential to support improvements in the capacity of teachers to deliver instruction that would enable all children to achieve to high academic standards, as well as to respond to the growing public concern that there be some link between teacher salaries and teacher performance. This innovation, knowledge and skill-based pay, rewards teachers with base pay increases and/or bonuses for acquiring and demonstrating specific knowledge and skills needed to meet educational goals, such as improving student achievement. The application of this pay concept to K-12 education has been suggested by Conley and Odden (1995), Mohrman, Morhman and Odden (1996), and Odden and Kelley (1997). This article reports on a study of seven knowledge and skill-based pay systems for teachers that have been developed by U.S. schools or districts.

Knowledge and skill-based pay can be better understood by contrasting it with two other teacher pay systems. Unlike the traditional single salary schedule, on which teachers progress through the salary schedule based on the number of years of service and the additional degrees or college credits they acquire, knowledge and skill-based systems provide pay increases when teachers demonstrate, usually though some form of performance assessment, that they have acquired and can apply classroom-relevant knowledge and skills. Ideally, pay progression is based on mastering a sequence or of knowledge and skills that represent higher levels of expertise or higher levels of teaching practice. The intent of knowledge and skill- based pay is to supplement or replace the traditional schedule with a pay system that motivates teachers to acquire and demonstrate the application of knowledge and skills that more directly contribute to better school performance and student achievement. The importance of seniority as a basis for pay is reduced or even eliminated.

The other contrast is with merit pay programs. Merit pay typically involves providing individual teachers with base pay increases by allotting a fixed fund of money based on administrators' subjective judgments of teacher performance during the prior year. While knowledge and skill-based pay programs also reward individual teachers, the reward is based on demonstrating knowledge and skills with respect to public, relatively detailed standards or descriptions of practice. These standards both guide assessor judgments and make known to teachers 'up front' what they need to do to demonstrate the knowledge and skills. Since any teacher who demonstrates the skills receives the reward, teachers do not compete for a share of a fixed fund or merit pay pool. These features of knowledge and skill-based pay may make it more effective in motivating more teachers than merit pay.

Because knowledge and skill-based pay programs are new and quite rare in the K-12 sector, it is not yet possible to obtain definitive evidence about the success of these programs in influencing instructional capacity or in improving student achievement. This article therefore concentrates on describing and comparing seven pioneer knowledge and skill-based pay programs. To do so, a set of dimensions were derived from an explicit theory of action which links knowledge and skill-based pay to improvements in instructional capacity and student achievement, and from the literature on knowledge and skill- based pay in the private sector.

The Theory of Action for Knowledge and Skill-Based Pay

Knowledge and skill-based pay systems have the potential to positively impact instructional capacity, and in turn student achievement, in three ways. First, they provide incentives for teachers to develop specific knowledge and skills needed to increase instructional capacity. More highly skilled teachers, in turn, have the capacity to deliver higher quality instruction, which, when combined with motivation to improve instruction and a context conducive to applying the skills, should lead to improved instruction. Second, by allocating higher pay to teachers who have these skills, these programs should help attract and retain high capacity teachers, and by denying higher pay to teachers without the skills, discourage lower capacity teachers from staying. Over time, the average skill level of a faculty should increase, improving the average quality of instruction. Third, a well-developed knowledge and skill-based pay systems rests on a model of competence that can also be used in teacher evaluation, professional development, and even recruitment and selection. To the extent this model informs these human resource management functions, the organization communicates and reinforces a normative vision of quality instruction. This model can also be used by teachers as a guide to professional development activities, a framework for self-reflection and self-evaluation, and a vocabulary for the discussion of teaching practice. Over time a shared conception of quality instruction should develop that supports teacher skill seeking and efforts to improve practice. This in turn contributes to improved student achievement. Figure 1 below summarizes this "theory of action" for knowledge and skill-based pay.

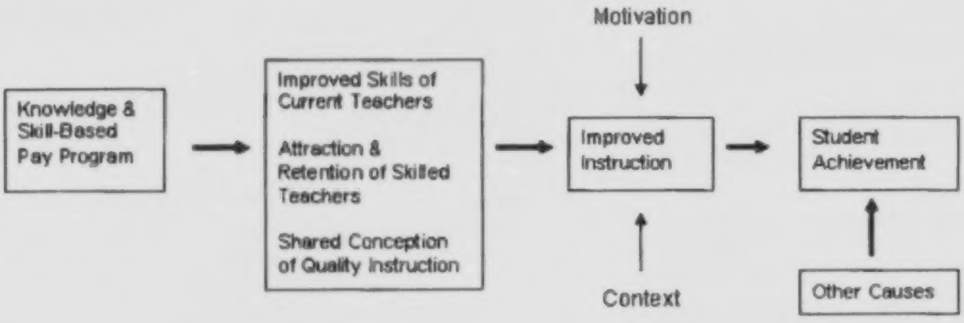

Motivation
Knowledge & Skill-Based Pay Program
Improved Skills of Current Teachers
Attraction & Retention of Skilled Teachers
Shared Conception of Quality Instruction
Improved Instruction
Student Achievement
Context
Other Causes

Figure 1. Theory of Action for Knowledge and Skill-Based Pay

The most important process by which knowledge and skill-based pay is expected to function to improve instructional capacity is by providing a pay incentive for knowledge and skill acquisition. However, simply offering teachers a pay increase or bonus will not necessarily motivate them to acquire the needed skills. We have used a modified version of Expectancy Theory (Vroom, 1964) to develop a model to identify what a knowledge and skill based pay program needs to do in order to motivate skill acquisition (see Figure 2).

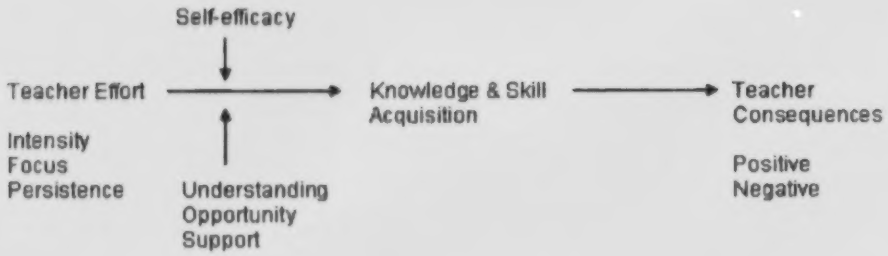


Figure 2. Motivational Model for Knowledge and Skill-Based Pay Based on Expectancy Theory

This model suggests that in order for knowledge and skill-based pay to motivate effort toward skill acquisition, teachers must first believe that it is likely that if they put forth the effort, they can actually acquire the specified knowledge and skills. This is called the expectancy perception, and is symbolized by the arrow running from effort to knowledge and skill acquisition in Figure 2. This perception is influenced by several factors, including the teacher's sense of self-efficacy for acquiring the skills and conditions the organization can more easily influence, including the degree to which the teacher understands what knowledge and skills are required and how they are to be demonstrated, the perceived degree of peer and administrator support for developing the skills, and the perceived availability of opportunities to develop the skills (such as high-quality professional development). To the extent that the teacher understands the skill requirements, believes that peers and administrators support their acquisition, and believes there are the required opportunities to develop the skills available, s/he will be more likely to believe that if s/he tries, s/he will be able to acquire the skills.

Teachers must also believe that there is a strong connection between acquiring the skills and positive consequences such as receiving the pay increase. This link is called the instrumentality perception, and it reflects the common-sense idea that if teachers do not believe that the reward is contingent on acquiring the skills, then the promised reward won't motivate skill-seeking. This link is represented by the arrow from knowledge and skill acquisition to consequences in Figure 2. In order for this perception to be strong, teachers must believe that the promised pay increases will be provided when the skills are demonstrated, and will not be provided when they are not. One set of conditions likely to support this belief include a reliable source of funding for the pay increases and the past performance of the organization in keeping promises to teachers. Another condition is that the methods used to assess knowledge and skill acquisition be fair, valid, and reliable. If teachers believe that favoritism or measurement error determines how well one does on the assessment, rather than their true skill level, they will be less likely to expend effort to acquire the skills. If skill acquisition cannot be validly measured, pay increases will be less contingent on skill acquisition, and when teachers realize this, they will be less motivated to acquire the skills.

Acquiring and demonstrating the skills must also have consequences teachers value. While it is safe to assume teachers value pay increases, these rewards also must be large enough to be perceived as worth the effort expended to acquire the specified skills. These rewards will also be more motivating if the knowledge and skill model on which the program is based is accepted by teachers as consistent with their conceptions of quality instruction and a highly-skilled teacher. Presumably, most teachers want to consider themselves good at what they do and are interested in developing their skills toward their ideal of a highly-skilled teacher. They may find this process of development intrinsically rewarding. If the knowledge and skill model is contrary to this ideal, teachers are presented with a choice: develop different skills and get more pay, or develop skills consistent with the ideal and forgo the extra pay. The extrinsic and intrinsic rewards work against each other. It is likely that the extrinsic pay reward will have a more motivating effect if it is consistent with the intrinsic reward. This means that the knowledge and skill model needs to be consistent with teachers' beliefs about what constitutes of a highly-skilled teacher. Finally, teachers may also value avoiding certain negative consequences, such as not being recognized as highly- skilled or expert. Avoiding these may also be motivating, especially if the definition of "expert" is shared by school-level peers.

Comparison Dimensions

Based on the theory of action, the motivational model, and the research and practitioner literature on private sector knowledge and skill-based pay systems, seven dimensions were developed to structure the analysis and comparison of the seven cases of knowledge and skill- based pay we studied.

1. Impetus Or Motivation For Developing The Knowledge And Skill- Based Pay Program

The theory of action assumes that policy makers choose to initiate these programs in order to improve instruction and in turn to improve student achievement. Alternatively, adoption of new forms of teacher compensation by pioneer organizations may be motivated by the desire to appear innovative or by the desire of influential decision makers to implement strongly-held ideas about which teachers should be paid more. In addition, a pay system change can present an opportunity to further other agendas, such as providing additional pay for all teachers or assuring the public that teacher pay is related to teacher performance. The motivation for moving to knowledge and skill-based pay is important, because it is likely to be related to design features such as the knowledge and skills in the model and the extent to which the new pay structure departs from the traditional salary schedule. One might expect that where the primary motivation is to improve student achievement, the knowledge and skill model will focus on instruction, and pay increases for developing instructional skills will be greater.

2. The Design Process

The motivation model suggests that teachers' views of the fairness of various aspects of the program and their acceptance of the model of good teaching implied by the knowledge and skills rewarded will influence their motivation to acquire the knowledge and skills. One way to promote the perceived fairness and acceptability of the system is to have teachers participate in its design. The private sector prescriptive literature on compensation program design (e.g. Lawler, 2000; Wilson and Phalon, 1996; Ledford, 1989) has advocated such employee participation. So did Odden and Kelley (1997) for education. Participation is thought to increase the level of information employees have about the program's rationale and operation. Employees also have valuable information to share about what they value, how the program is likely to work in practice, and how they are likely to react to it. Participation is also thought to increase "buy- in". Because a high level of participation is likely to result in greater acceptability and perceived fairness, an important facet of this dimension is the degree to which teachers participated in the design of the program.

One form of teacher participation is through collective bargaining. However, it may be difficult to design a knowledge and skill-based pay program through the standard adversarial collective bargaining process of proposal and counter-proposal, with each side seeking maximum advantage. Knowledge and skill-based pay programs require a coherent design based on some agreed-upon conception of good teaching. Many technical details, such as how knowledge and skills will be assessed, need to be addressed. So it is expected that these programs would be designed either outside the formal contract negotiation process or though an interest-based process (Fisher and Ury, 1981) that focuses the parties' attention on mutual goals.

Another important aspect of the design process is how program designers decide what knowledge and skills to reward. Designers in the private sector appear to have used inductive, deductive, or adaptive approaches. The inductive approach involves using job analysis or relying on research to identify those knowledge and skills likely to contribute to employee performance. One version of this method is to study known good and average performers to find out what knowledge and skills differ between these groups (Spencer and Spencer, 1993, American Compensation Association, 1996). The deductive approach involves starting from the organization's strategy, then trying to identify the knowledge and skill employees need to carry it out (Heneman and Thomas, 1997, American Compensation Association, 1996). The adaptive approach involves starting with a knowledge and skill model developed elsewhere, then changing it to fit local goals and conditions. Though the use of the adaptive method in the private sector has been criticized because it does not provide a unique source of competitive advantage (Ledford and Heneman, 2000; Zingheim, Ledford, and Schuster, 1996), it avoids 'reinventing the wheel', especially for those core knowledge and skills likely to be common across organizations.

In the K-12 sector the core technology of instruction is similar across schools. Since there is currently little competition among schools, there is little incentive for very different specifications of knowledge and skills to be identified. There are also economies of effort to be realized by adapting work already done by recognized bodies of experts, such as the standards proposed by the Interstate New Teacher Assessment and Support Consortium (1992), state teacher licensing standards, the National Board for Professional Teaching Standards (1999)(NBPTS) standards, or the Danielson's Framework for Teaching (Danielson, 1996). Thus we might expect that many of these pioneer knowledge and skill-based pay programs would have adapted external standards, perhaps adding locally-important skills or modifying language to fit local conditions, rather than attempting to develop an organization-specific model. This approach also allows program designers to appeal to the authority of these external experts when seeking support from teachers and the community.

3. Types and Structure of Knowledge And Skills Rewarded.

At the heart of a knowledge and skill-based pay program is the specification of the knowledge and skills teachers will be rewarded for developing. The theory of action assumes that the knowledge and skills specified will be those teachers need to deliver instruction that contributes directly to student achievement. Thus an important facet of this dimension is the degree to which the knowledge and skills rewarded are related to instruction.

Another important facet is the extent to which the knowledge and skills rewarded are organized into an integrated model with a defined continuum of skills or expertise. Knowledge and skill-based pay programs in the private sector often structure the knowledge and skills rewarded into a set of career levels (Jones, 1995; Daniels, 1997), levels defined by rating scales (Heneman and Thomas, 1997; Gorsline, 1996; American Compensation Association, 1996) or sequences of skills to be mastered (Gupta et al, 1986; Jenkins et al, 1992). Odden (2000) and Odden and Kelley (1997) sketched a number of different structures of knowledge and skills representing a progression from entry level to accomplished teaching. Such a structure could provide a roadmap for teachers seeking to develop their knowledge and skills as well as convenient attachment points for pay increases. It could also be used to align other parts of the human resource management system, especially professional development programs, and as a guide for teachers working to develop mastery of quality instruction.

4. How Knowledge And Skill Acquisition Is Assessed.

The motivational model suggests that knowledge and skills should be assessed in a way that teachers see as fair and valid, and the theory of action implies that the assessment method must ensure teachers can apply the skills in practice. The traditional degrees and credits seem to be viewed as fair by teachers, but they may not have high validity as indicators of whether skills can be applied in the classroom. Properly constructed and administered, performance-based assessments, which function as samples of teachers' instruction, have the potential to ensure that the skills can be applied and to be perceived as valid and fair, due to their close connection with practice. Thus one facet of this dimension is the extent to which performance-based assessments are used, rather than degrees and credits, to provide evidence of knowledge and skill acquisition.

While private sector knowledge and skill-based pay programs typically appear to depend on relatively simple, locally-developed assessments (Heneman and Ledford, 1998), program designers in the K-12 sector have the option of using externally-developed assessments, such as PRAXIS III (Dwyer, 1998), the Framework for Teaching (Danielson, 1996), or the NBPTS assessments. Thirty-one states and more than 200 districts provide some salary incentive for certification. (National Board, 2001). This avoids the expense and effort of developing local assessments for core teacher skills that are likely to be common across districts or schools (Milanowski, Odden, and Youngs, 1998, Heneman and Ledford, 1998). External assessments may also have the potential for greater validity and fairness than assessment developed locally, due to the greater expertise and resources of their developers. External assessments could be used in combination with local assessments to maintain the rigor of the system. The teacher performance evaluation literature (e.g. Wise et al, 1984) suggests that local assessors such as principals face many incentives to be less than rigorous. If almost all teachers are judged to have the skills, due to leniency of local assessors, the contingency between skill acquisition and receiving the reward the motivational model postulates as necessary is reduced. (The reward won't motivate effort toward skill acquisition if the assessors certify teachers without the skills as eligible for the reward.) Odden (2000) outlined a model knowledge and skill-based pay structure that combined the use of external and local assessments. So a second facet of this dimension is the extent to which external and locally-developed assessments are used to provide evidence of knowledge and skill acquisition.

5. The Size and Structure of the Knowledge and Skill Incentives.

The theory of action proposes that the extra pay offered will motivate teachers to acquire the knowledge and skills needed to improve instruction. To motivate, the pay incentive provided must be valued. Experience with the traditional salary schedule suggests that teachers value pay rewards enough to collect years of seniority, credits, and degrees. But to motivate the acquisition of the new, possibly hard-to- master skills needed to improve instruction, the incentives must be of sufficient size to attract teachers' attention and to be perceived as commensurate with the effort needed to acquire the skills. It is reasonable to expect that the greater the size of the incentive, the more motivational effect, all else equal. So an important facet of this dimension is the size of the incentive offered.

To the extent that knowledge and skill rewards replace the traditional pay increases for seniority and educational attainment, we might expect teachers to be more motivated to attain the skills, since the traditional opportunities for pay increases have been reduced. A more radical change in the pay structure, de-emphasizing seniority and educational attainment unrelated to classroom instruction, sends a stronger signal that new knowledge and skills are needed. Knowledge and skill pay programs might be located on a continuum ranging from those that supplement the traditional salary schedule by simply adding additional pay opportunities based on knowledge and skill acquisition, to complete replacement of the traditional schedule's seniority steps and educational attainment lanes with a set of pay levels based only on knowledge and skill attainment. One might expect that the greater the perceived need to improve instruction, the more the traditional salary schedule would be modified and the larger the incentives for knowledge and skill acquisition would be.

A knowledge and skill-based pay system carries risks for teachers accustomed to automatic pay increases based on seniority. It may be particularly unattractive to more senior teachers because it places less emphasis on seniority as a criterion for pay differentiation, and can require developing new skills, which may not be as good an investment of effort for them. To get a knowledge and skill-based pay program accepted may require some provision that compensates teachers in some way for the increased risk or reduces the threat that the emphasis on new skills can represent to senior teachers. Thus another feature of interest is whether the programs include provisions intended to make the new system acceptable to potential opponents such as veteran teachers.

6. Alignment of Other Human Resource Programs in Support of the Knowledge and Skill Model

Ensuring that the professional development programs available to teachers are aligned with the knowledge and skill model is likely to be a determinant of program success, because according to the motivational model, teachers need to perceive that opportunities to acquire the rewarded skills are available in order to believe that their efforts are likely to be successful. Private sector employers appear to take on the responsibility for providing and communicating opportunities to acquire skills, to ensure availability and to show employees their efforts to acquire skills are being supported (American Compensation Association, 1996; Jenkins et al, 1992). Thus an important of aspect of alignment is whether organizations provide professional development opportunities linked to the knowledge and skills their pay systems reward.

A knowledge and skill model can also provide a foundation for other human resource management programs such as performance evaluation, recruitment, and selection (Spencer and Spencer, 1993, Schippman et al, 2000). The theory of action postulates that a human resource management program aligned with the model will contribute to the development of a shared conception of good instruction consistent with the model. If the model is shared with job candidates during recruitment, those who do not believe that they can develop the skills or are not in agreement with the underlying philosophy of instruction may "self-select" out of the hiring process. Selecting new teachers based on the knowledge and skill model helps to ensure those who are hired have the skills, or the potential to develop them. If teachers select the district or school and the district or school selects teachers based on the model, convergence on the conception of instruction it embodies should increase over time. With respect to current staff, if teacher evaluation is made consistent with the knowledge and skill model, this will avoid confusing teachers about what it values as good teaching, and teachers will not be faced with two unrelated assessments on which they must spend time and energy. Teachers should be more likely to use the model to guide their own professional development efforts, and to absorb the model as the appropriate way to think about teaching, again reinforcing a share conception of instruction. Thus a second aspect of alignment is the extent to which the knowledge and skill model is integrated with other human resource management programs besides pay and professional development.

7. Costs and Funding

While knowledge and skill-based pay offers substantial benefits, it is also likely to require additional investments, including the costs of increased professional development and additional administrative overhead (e.g. assessment and record-keeping) as well as of higher salaries. Private sector experience with knowledge and skill-based pay programs suggests that administrative costs increase and per-employee salary costs increase. (Gupta et al, 1986; Jenkins et al, 1992). Individual pay increases are thought to be offset by increases in productivity and greater flexibility in staff utilization due to cross-training. However, these offsets are less likely to appear in the K-12 sector because increased productivity, in the form of higher student achievement, typically does not allow reductions in staff nor savings in materials or equipment. (Nor is it immediately marketable for increased revenue.) The knowledge and skills are not those that allow teachers to do more different jobs, therefore allowing elimination of support staff. Therefore in the long run we would expect higher costs, which need to be funded by new money or reallocation of existing resources.

Case Selection, Data, and Method

The cases compared here include of six school districts and one charter school all of which had adopted some form of knowledge and skill-based pay. The cases were selected based on project researchers' knowledge of districts or schools designing and implementing these pay programs, and a survey of state department of education and teacher association staffs which asked them to identify districts with innovative pay systems. These "early adopters" are not representative of U.S. schools or districts, merely illustrative of the variety of knowledge and skill- based pay programs that are being developed and of the process of design and implementation. Description and comparison is based on the programs' operation or design as of the 1999-200 school year. In each case, project staff visited the district or school, in some cases multiple times, during the 1998-2000 period. Administrators, union officials, and in some instances, teachers were interviewed. A semi-structured interview protocol guided most of the interviews. Documents describing the program were also collected, and in some cases internal research done by the districts to evaluate the programs was obtained. The researcher who visited the site wrote a case description from which the information relevant to the comparison dimensions was abstracted. The Appendix provides a brief description of each case site. Extended case descriptions are available at www.wcer.wisc.edu/cpre. In a few cases, additional contacts were made by the author to clarify information in the case descriptions. The author then summarized the features of each case along the comparison dimensions, then attempted to identify patterns and important differences, and to draw conclusions about the implications of the experiences of these early adopters for research and program design.

Program Comparisons

The similarities and differences among the seven programs are presented below, structured according to the seven comparison dimensions.

Motivation For Developing The Knowledge And Skill-Based Pay Program

The varied motivations for pay system change we found suggest that knowledge and skill-based pay was not simply seen as a way to improve student achievement through improving the skill level of current staff, as assumed by our theory of action. Though supporting improved instruction was a common goal, there were other important reasons for initiating change. It does not appear that most of these early adopters were primarily focused on using the programs increase teachers' instructional capacity in order to improve student performance. Decision-makers at most of the sites did not appear to have based their programs on a theory of action like the one described above. Table 1 summarizes the key factors in each case.

Table 1
Initiation of Knowledge and Skill-Based Pay Design Process

Site	Primary Motivation for Developing KSBP	Supporting State Policy Initiatives	Champion(s)	Labor-Management Relations
Cincinnati	Strategic planning process identified KSBP as one district strategy for improving student achievement.	State proficiency tests Change in State licensingstandards	Union Bargaining Chair, Associate Superintendent, outside consultant from university.	Variable; but underlain by trust relationship between key union and management staff
Coventry	Desire to differentiate pay according to performance and support new model of instruction. Secondarily, concern about	None directly, though new state funding formula may have provided some of the	Superintendent, union president	Cooperative, after period of conflict in the 70's

	pay inequities between junior and senior teachers; desire to keep good teachers in the classroom rather than moving to administration to receive more money	additional funds needed.		
Douglas County	Part of broader compensation redesign aimed at reassuring public concerned about accountability for use of public funds	State teacher licensing reforms	Assistant Superintendent for HR; union president	Cooperative
Limon	Desire to link pay with teacher performance coupled with teacher dissatisfaction with prior individual pay for performance system. Desire to add group reward component linked to state assessment results.	State financial incentives for local districts to adopt pay for performance systems; state assessment system.	Superintendent initially, then teachers	Basically cooperative, due to new superintendent and relative weakness of union in "right to work" state.
Manitowoc	Desire to provide incentives for teachers to develop their skills in order to implement more constructivist instruction; improve teacher retention; support union-initiated professional development program	State licensing reforms	Superintendent initially; joined by regional union representative	Cooperative; new leaders changed previous more adversarial relationship
Robbinsdale	Concerns with recruiting and retaining quality teachers in competition with other local districts; concern about pay inequities between junior and senior teachers.	None directly.	Union president	Cooperative, based on stable district and union leadership
Vaughn	Part of broader compensation redesign aimed at improving recruitment and retention of good teachers, desire to add teacher accountability to external accountability provided by charter, and to address perceived pay inequities between junior and senior teachers.	Charter status and performance pressure from explicit charter goals	Principal, initially, then junior teacher and several more senior staff	No union; relatively high level of trust between teachers and administrators based on commitment to charter, participatory governance structure, and charisma of principal

In Cincinnati and Vaughn, cases where external accountability pressures emphasized the need to improve student achievement, the primary rationale for the programs was not expressed in terms of remedying a knowledge or skill deficit among current staff. Rather, the programs seemed to be intended to motivate staff to change practice and to reward more accomplished teaching. In Cincinnati, this was supplemented by dissatisfaction with the current teacher evaluation system and changes in the state licensing system. At Vaughn, the recruitment and retention of highly-skilled teachers was an important additional aim. In Robbinsdale, recognizing and rewarding accomplished teachers, and recruitment of skilled teachers, appear to have been the major goals. In Coventry, program designers wanted to differentiate teacher pay according to performance and to keep good teachers in the classroom as well as to support a particular vision of quality instruction. In Manitowoc, the superintendent's vision of quality instruction and his desire to provide incentives for teacher learning were joined by the union leadership's interest in improving pay while staying under state-imposed expenditure limits and supporting a professional development initiative developed by the state teachers' association. Both the superintendent and the association leadership wanted to improve retention, and begin adapting the pay system to state licensing changes. In Douglas County, knowledge and skill-based pay came about as part of a pay system redesign primarily intended to respond to public pressure to link teacher pay and teacher performance, in order to improve accountability for the use of public funds. Limon, the current program replaced one in which pay increases were based on individual teacher evaluations. Program designers there found a way to respond to public interest in linking pay to performance and teachers' concerns about unfairness of the old system by rewarding both individual professional development and meeting building and grade-level student achievement goals. Perceived inequities in the traditional salary schedule between younger, high performing teachers and more senior teachers contributed to motivating pay system change in Coventry, Robbinsdale and Vaughn. This is interesting given that one advantage often cited for the traditional salary schedule is that teachers perceive it to be highly equitable (Odden and Kelley, 1997).

State policy, while not a primary driver of pay system change, was an important background condition in many of the cases. Teacher licensing policy provided a reinforcement for change in Cincinnati and Manitowoc. State incentives may have provided motivation for the Limon's initial experimentation with non-traditional pay systems, and for the district to continue to include teacher and student performance elements in its current plan. But while all of the sites were in states with some form of student testing and accountability program, only the two sites with relatively low student achievement (Cincinnati and Vaughn) felt much pressure from these programs. In the others, student achievement was either relatively high or not a major issue in the other communities.

While in all of the cases, a champion or set of champions was important in keeping it going to a successful conclusion, in three the champion's own agenda was a key impetus to initiating change. In Coventry and Manitowoc, the programs were initiated partly to pursue the superintendent's personal vision of good instruction, though in Coventry the union president actually got discussion going by proposing rewards for National Board certification. In Robbinsdale, the former union president initiated discussions with the district based on his desire to ensure that new teachers who fit his conception of a good teacher would be available when it became necessary to replace retirees. At Vaughn, the principal began to explore pay innovations to strengthen teachers' sense of accountability for student performance by adding individual stakes to the overall external accountability provided in the charter. It is interesting that in four of the seven cases, a union official was one of the champions, and in a fifth, a Uniserve representative was a key catalyst of innovation. This suggests that teachers' unions can be supportive of changing the traditional salary structure. Where union and management relations are good, and a high- trust relationship exits between union leaders and at least some management leaders, it appears that teacher compensation innovation can be successfully initiated.

Process Used To Design The Program

Table 2 summarizes three key aspects of the design process at these sites: the type and level of teacher participation in design, the relationship to the collective bargaining process, and the methods used to identify the knowledge and skills to be rewarded.

Table 2
Characteristics of Knowledge and Skill-based Pay Design Process

Site	Relationship to the Collective Bargaining Process	Type/Level of Teacher Participation	Method of Knowledge & Skill Identification
Cincinnati	Commitment to develop plan agreed to in contract. Design process took place outside normal collective bargaining though system of union-management committees. Steering committee jointly chaired by union and management representatives. Teachers voted to approve contract including the results, and have opportunity to vote out system before 9/2002 implementation.	24 teachers from a variety of schools participated on the various committees.	Adaptation of standards for teacher performance found in the Framework for Teaching (Danielson, 1996)
Coventry	Worked out as part of interest-based bargaining process	Limited to members of bargaining team, though rank and file teachers participated in design of the teacher evaluation system incorporated into the plan.	Adaptation of NBPTS standards for one pay incentive; inductive process drew from best practice literature (including work of T. Sizer and F. Newman) & district action research for the other
Douglas County	Commitment to develop plan agreed to in contract. Design process took place outside normal bargaining via a 30 member performance pay committee.	Performance pay committee included 20 teachers from a cross section of union members	Deductive and inductive processes used to develop skill blocks; adaptation of NBPTS and Colorado licensing standards for the outstanding teacher award.
Limon	No formal contract. Program concept developed by superintendent and teacher representatives as part of informal negotiations. School Board passed proposal & teachers' association agreed to try proposal.	3 teachers worked with Supt. to develop concept; detail design done by a committee with 7 teachers and one administrator.	Left to teacher and building administrator, based on district-provided guidelines.
Manitowoc	Worked out as part of interest-based bargaining process and approved as part of teacher contract.	8 association bargaining team members participated as part of bargaining process	Inductive, based on education research; adoption of NBPTS standards.
Robbinsdale	Initial concept and outline of skill areas and pay levels agreed to in bargaining. Detail design by set of union-management committees.	A few teachers participated as members of the bargaining team. More teachers participated as members of each of 8 committees responsible for defining skill levels and methods of assessment.	Adoption of NBPTS standards; induction from research and experience.
Vaughn	No collective bargaining.	Design by small group of teachers and administrators, then extensive discussions via informational meetings in committees of governance structure.	Primarily deductive from educational goals in charter, with some adaptation of the Framework for Teaching for the rubrics

Teacher participation. Large scale teacher participation in the design process was present in the three largest organizations (Cincinnati, Douglas County, and Robbinsdale). In these districts a formal committee process was used to involve a substantial number of teachers in some aspect of system design. These cases suggest that broad teacher involvement can have a substantial influence on the content of the plan. For example, in Cincinnati, teachers on the committees probed for ambiguities in the design proposals and contributed specific ideas for design, as well as pushed for provisions to reassure teachers about fairness. The input of National Board certified teachers was influential in persuading other teachers on the committee to take the risk of trying a new system of evaluation and pay. Of the smaller organizations, Limon provided for relatively extensive participation given the size of the district, while in Coventry and Manitowoc teacher participation appeared limited to the association bargaining team. A relatively small group developed the Vaughn plan, though it was adopted by a vote of the governance committee on which teachers were heavily represented and after considerable formal and informal discussion among the faculty. However, involving a substantial number of teachers does not guarantee broad communication. Outside evaluations of both Douglas County (Hall and Caffellera, 1997) and Cincinnati (Milanowski and Kellor, 2000b) suggest that many teachers who had not been active participants in the design process did not seem informed on some aspects of the systems. In Cincinnati, the large size of the district and the complexity of the program seemed to require more intensive district or association-sponsored communications efforts than were initially undertaken.

Relationship to collective bargaining. As expected, none of the programs were developed though traditional adversarial collective bargaining. It was also expected that knowledge and skill-based pay programs would be too complex and time- consuming to work out in detail through the normal negotiation process. However, in three of the cases, the details of the process were worked out within the negotiation process. It appears that pay changes of substantial complexity can be developed in the bargaining process, as long as the parties have achieved a high level of trust and focus on a vision or desired goal shared by both sides. Where the programs' details were bargained, a clear sense emerges from the case studies that the shared vision or goal was an important influence in keeping the discussions from getting sidetracked by issues of who gains and who loses from particular details of the program.

Knowledge and skill identification. There was no one method of knowledge and skill identification that dominated in these cases. As expected, many programs made use of existing standards or definitions of good teaching. In five of the seven cases, an external set of teacher standards, either the NBPTS or the Framework for Teaching (Danielson, 1996) were influential. Cincinnati is the clearest example of adaptation. Starting with the Framework for Teaching, the design committee examined each component and revised wording to fit the district context. Adapting the Framework for teaching allowed the district to design a system in a relatively short time. In contrast, Robbinsdale, though beginning with NBPTS standards and making Board certification a major determinant of knowledge and skill-based pay increases, had not been able to implement its system in the school year intended in part because of the difficulty in defining the key indicators of knowledge and skill in the parts of its system not related to the NBPTS standards. In Coventry, the Framework for Teaching is the basis for the teacher evaluation system, but the pay incentives are based on separate standards. One provision is based on the NBPTS certification, and the other on locally-developed criteria with content that differs from both Framework and the Board's standards. The programs in Douglas County, Limon, Manitowoc, and Vaughn were not primarily based on existing external standards, though Vaughn did adapt the format of the Framework for Teaching's rubrics. The strengths and weaknesses of the deductive approach are illustrated by the Vaughn case. The knowledge and skills developed were closely tied to school goals, so they had the potential to focus all teachers on key skills. However, the process of identifying the skills and the standards for measuring them was difficult and time consuming. As a consequence, in the first year the criteria for knowledge and skill demonstration were not well specified, causing many of the initial participants to have concerns about fairness of application (Milanowski and Kellor, 1999). These experiences suggest that adapting an existing model of teacher practice may be the most efficient way to get a knowledge and skill-based pay system up and running.

Knowledge and Skills Rewarded and Their Organization into a Structure

Table 3 summarizes the knowledge and skills rewarded in the cases, and how (if at all) the knowledge and skills were organized into some form of developmental sequence or set of performance levels. As expected all programs rewarded knowledge and skills relevant to instruction, especially pedagogical skills. There are differences in emphasis, however. Coventry, Douglas County's skill- blocks, and Manitowoc appeared to be trying to promote constructivist or "authentic" instruction, while Cincinnati and Vaughn were concerned with a more generic model of good teaching, though with some constructivist elements. Limon allowed the teacher and/or building administrator to determine what sort of instructional skills should be developed within broad district guidelines. Robbinsdale had not yet worked out its model in detail at the time of our study. None of the programs appeared to emphasize mastery of content-specific pedagogy (Shulman, 1987, National Commission on Teaching and America's Future, 1996), except as embodied in NBPTS certification, though some of Cincinnati's and Coventry's locally-developed rubrics referenced it and some of Vaughn's represented basic aspects of it.

Table 3
Knowledge and Skills Rewarded—Content and Structure

Site	Knowledge and Skill Domains Identified	Developmental Levels of Knowledge and Skills
Cincinnati	Three part system: 1) 4 core domains: planning and preparing for student learning, creating an environment for learning, teaching for learning, and professionalism; specific behavioral standards in each domain with rubrics describing 4 levels of performance on each standard 2) content knowledge; 3) NBPTS standards(a).	Core of system had five developmental or career levels; aggregation of rubric scores on standards define career level.
Coventry	Two separate pay provisions. One used NBPTS standards(a); The other (RHODE program) covered authentic pedagogy (instruction and assessment), self-reflection, differentiating instruction, family and community involvement, and professional development.	Two separate programs with limited overlap; they did not represent a developmental sequence, though RHODE could be useful in preparing for NBPTS certification.
Douglas County	Two part system: 1) 9 skill blocks covering technology, authentic assessment, and diversity; 2) Outstanding Teacher award with options using: a) NBPTS standards; b) standards-based instruction; or c) assessment and instruction, content and pedagogy, and collaboration. Standards for Outstanding Teacher were a mixture of knowledge and skill descriptions and descriptions of behavior.	Two separate programs with limited overlap; some of the skill blocks represent developmental sequences.
Limon	Program did not specify knowledge and skills to be sought, leaving this up to teacher & administrator based on general guidelines that emphasized the need to focus on instruction and student learning.	Unstructured; content and sequence open to development by teacher and administrator
Manitowoc	Three aspects of system: 1) district-developed courses covering authentic instruction, technology, writing instruction; 2) NBPTS standards (a); 3) content knowledge represented by degrees and credits. Knowledge and skills primarily defined in terms of courses and certifications	Program components are independent ; no developmental levels identified at time of study, though district courses and Professional Development Certificate could prepare teacher for Board certification.

Robbinsdale	NBPTS standards (a), content knowledge, classroom teaching, program/curriculum design, district and school leadership, parental/student satisfaction. Knowledge and skills defined mostly in terms of indicators such as NBPTS certification and documentable teacher accomplishments.	Program had ten independent elements that are evaluated, and the evaluations aggregated to produce a pay level. The different elements represented multiple ways to define good teaching rather than a developmental sequence.
Vaughn	Core system based on 11 locally-defined domains of skill in lesson planning and classroom management, literacy, language development, technology, special education inclusion, mathematics, history and social science, and science pedagogy, instruction in primary language for English learners, arts. Additional knowledge and skills rewarded defined by NBPTS standards (a), Masters' degree, state licensure level.	Core of system defined 3 levels for additional pay: level 1 based on achieving an average rubric score of 2.5 in 6 of the "essential" domains; level 2 required an average of 3 in those domains, then provides additional pay for rubric score of 3 in any of 5 additional domains; level 3 based on achieving an average rubric score of 3.5 in all domains.

Note: (a) The content of the National Board standards varies by subject and level among the 30+ certifications offered; however, almost all standards include the domains of knowledge of students, knowledge of subject, knowledge of pedagogy, creating a learning environment, use of a variety of assessment methods, reflection on practice, and collaboration with parents and colleagues.

Most of the programs are eclectic in the way they specify what knowledge and skills rewarded. While most of the rewards in the Cincinnati and Vaughn programs are based on developing knowledge and skills that are described in terms of teaching behaviors or skilled performance, there is some reward provided for degrees or certifications analogous to the credits in the traditional schedule. Another set of programs, Douglas County, Coventry, and Robbinsdale, mixed external certifications with more or less detailed descriptions of desired performance or behavior. Limon provided relatively little guidance, leaving the teacher and administrator wide leeway as to the knowledge and skills to be developed. The Manitowoc program specified its knowledge and skills in terms of courses and certifications, analogous to degrees and credits, rather than describing behaviors or skills.

In only a few of the programs, most notably those of Cincinnati and Vaughn, did the knowledge and skills specified approach the ideal of an integrated developmental sequence or structure of levels. The other programs had not organized the knowledge and skills into a core set of standards, nor provided a continuum of skill development that unified the knowledge and skill domains along a developmental path or career progression. Even the Cincinnati and Vaughn programs did not appear to integrate their local standards with the National Board standards, treating Board certification as an additional credential like a Masters' degree rather than as another developmental level. Though the programs in Coventry and Manitowoc were informed by a coherent vision of instruction on the part of their original champions, these programs did not include a developmental progression linked with pay increases at the time we studied them.

How Knowledge and Skill Acquisition Was Assessed

Table 4 summarizes the assessment methods used in each of the seven programs, including the use of external assessments. All use some form of performance assessment, though the extent to which these assessments are central to the program varies. The Cincinnati and Vaughn assessment systems were primarily based on demonstrating knowledge and skills via classroom performance and are part of the regular teacher evaluation. Programs that use National Board certification as a criteria for pay increases (Manitowoc, Robbinsdale, and Coventry) incorporated the performance emphasis of the Board's assessments. Coventry also used a performance-based approach in its local assessments. The assessments at the end of Douglas County's skill blocks are performance-based, though performance in training is not always the same as classroom performance. The guidelines for the Limon process emphasized connecting the professional development documented in the portfolio to classroom practice and student learning, though it is up to the teacher and administrator to implement these guidelines. One of Robbinsdale's performance dimensions involved principal evaluation via classroom observation, and several others are based on real-world accomplishments rather than degrees or credits. The observations were to be part of the regular teacher evaluation process, and the results one element in a teacher portfolio that documents knowledge and skill. Manitowoc's program mostly relied on indirect evidence like certifications and course attendance, more analogous to the traditional degrees and credits. Performance assessment was incorporated mostly though the incentive provided for National Board certification.

Table 4
Methods of Assessment

Site	Locally-Developed Assessments Used	External Assessments Used
Cincinnati	Internal assessment of performance in the four primary knowledge and skill domains by site administrators and peer evaluators with subject expertise. Types of evidence: 6 classroom observations; portfolio including artifacts such as lesson plans, student work, parent contact logs, professional development logs.	NBPTS assessment, degree completion, and licensure will be used to determine eligibility for additional pay elements.
Coventry	Internal assessment based on a portfolio prepared by teacher used for the RHODE program. Portfolio including evidence that/of: teachers know students, have prepared for and practiced differentiated learning, ability to motivate and support all students, family and community contact, and professional development. Also self- analysis of teaching and assessment	NBPTS assessment used for pay increment for NBPTS certification.

	of student work. RHODE: 9 element portfolio evaluated using rubrics. Each element scored for evidence of presence of 5 behaviors or outcomes each worth 1 point. Total score of 43 points qualifies for award.	
Douglas County	Performance-based assessment at end of each skill block done by course instructors. No specific rubrics or standards defined outstanding teacher; review of portfolio prepared for outstanding teacher award done by administrators.	NA
Limon	Portfolio documenting activities toward fulfilling professional growth goal evaluated by administrators. No specific rubrics or standards to evaluate skill acquisition	NA
Manitowoc	Mixture of external (NBPTS, Professional Development Certificate, degrees) and internal (local teacher-taught courses) opportunities. Standards or rubrics used to evaluate skill acquisition depended on course or certification.	Grades/degrees/certifications from higher education. NBPTS assessment used for pay element rewarding NBPTS certification.
Robbinsdale	Documentation of achievements via portfolio, classroom observations; student/parent surveys. Evidence evaluated by committee consisting of 3 appointees of superintendent, 3 appointees of union president. Specific rubrics/guidelines remained to be developed for most domains.	NBPTS assessment used for pay element rewarding NBPTS certification.
Vaughn	Classroom observations, artifacts such as lesson plans and student work evaluated by an administrator, grade-level peer, and self. Four-level rubrics used specific behavioral examples to define levels of performance in each domain.	Grades/degrees/certifications from higher education used for pay elements rewarding credentialing. NBPTS assessment used for pay increment for NBPTS certification.

Use of external assessments. Five of the seven programs included the NBPTS assessments. However, these assessments were not typically integrated with the local assessment system, nor used as measure of core teaching skills. In most cases, the NBPTS assessment was included because of pay incentives for NBPTS certification, which in turn was treated as an additional degree, rather than as an integral part of the knowledge and skill model. The exceptions are Coventry and Robbinsdale. In Coventry, while the domains measured by the local and Board assessments differ, the processes are similar, with the local process designed to help teachers prepare for the Board assessment. In Robbinsdale, the NBPTS assessment was the criterion for a sizable part of the knowledge and skill pay incentive, and seemed to form the conceptual anchor for the program, but there were several other locally-assessed ways for teachers to demonstrate knowledge and skill in order to increase their pay.

Validity and reliability. According to the theory of action, knowledge and skill-based pay programs require methods of assessment that are valid and reliable, and recognized as such by teachers. At the time of our studies, little information was available about the validity or reliability of the assessments in any of the programs. From the information available, it appears that the most common external assessments, those of the NBPTS, have at least as much reliability and validity as many accepted human resource selection and evaluation techniques (Milanowski, Odden, and Youngs, 1998; Jaeger, 1998). Only Cincinnati and Vaughn appeared to have confronted these issues with respect to their locally-developed assessments. Both have looked at the inter-rater agreement of their internal assessment systems and have been relatively satisfied that an acceptable level of agreement exists. At this point, no district has looked at the relationship between its assessments and independent measures of teacher performance, such as student achievement, though at least two were planning to explore this connection.

Information on teacher perceptions of the fairness of these assessments was available for three cases. In Cincinnati, teachers participating in the field test of the assessment system generally believed that the results of the process were fair, but many had concerns about the fairness of the process, especially with respect to administrator qualifications and the consistency of rating across administrators (Milanowski and Kellor, 2000b). At Vaughn, in the initial year, many of the teachers covered by the system had concerns about the consistency of the evaluations and the difficulties evaluators had in making the requisite number of classroom observations (Milanowski and Kellor, 1999). After the rubrics were more fully developed and problems with scheduling of observations addressed in the second year, fairness perceptions improved substantially (Milanowski and Kellor, 2000a). In Douglas County, while fairness perceptions were not explicitly assessed, Hall and Caffarella (1997) did interview and survey teachers about their reactions to the program. Teachers did not identify fairness concerns as a major issue, though several teachers did mention problems such as subjectivity in the evaluation of the teacher portfolios. It may be that the lack of expressed fairness concerns was due to the lower stakes of the assessments for individual teachers (due to the relatively small dollar amounts associated with the skill blocks and Outstanding Teacher award) the fact that individual teachers could choose to participate or not.

Size and Structure of Knowledge and Skill Incentives

Table 5 describes the programs' knowledge and skill-based pay structures. The programs can be roughly categorized as falling into three groups: those that have essentially replaced the traditional schedule (Cincinnati, Limon), those that have modified the schedule, typically by reducing the importance of degrees and credits in exchange for more performance-oriented representations of knowledge and skill (Vaughn, Robbinsdale), and those that have supplemented the traditional design by adding knowledge and skill-based elements (Coventry, Douglas County, Manitowoc). We expected that organizations feeling more pressure to improve student achievement would be more likely to modify or replace the traditional schedule, but the association is not that strong. Cincinnati and Vaughn fit the pattern of organizations under pressure implementing major pay change, and Coventry, Douglas County and

Manitowoc fit the pattern of less pressure and more incremental change. Limon was an exception, in that it replaced the traditional schedule completely, but for reasons other than providing a stronger incentive for knowledge and skill acquisition. Robbinsdale was another exception, with a relatively major pay system change but little pressure to improve student achievement. It should be noted that the potential impact of the more radical changes represented by the Robbinsdale and Vaughn pay systems were offset by limitations on who is covered by the new system. To allay apprehension on the part of senior teachers socialized to the traditional system, these two programs required that only new or less senior teachers participate, leaving others on the traditional schedule.

Table 5
Knowledge and Skill Pay Provisions and Relative Size of Incentive

Site	Pay Provisions	KSBP Incentive as % of Beginning Base and Maximum Salary	Provisions to Win TeacherAcceptance
Cincinnati	Core program: 5 career levels of teaching practice with salary ranges of $30,000, 32,000-35,750, 38,750-49,250, 52,500- 55,000, and 60,000-62,500; movement between levels based on knowledge & skill assessment, movement within levels through a limited number of steps based on seniority; additional base pay add-ons of $4,600 for Masters degree, 9,375 for Ph.D.; $1,250 for dual certification, $1,000 for NBPTS certification, up to $4,000 (time-limited) for skill blocks.	Base: 60.8% Maximum: 23.4%	High seniority teachers (22 years and up) can remain on old salary schedule or can volunteer to participate in new system.
Coventry	$6,500 add-on for life of NBPTS certification; $1,000 per year for four years based on achieving a cut-off score on a locally-assessed portfolio (RHODE program).	Base: 22.8% (19.8% for NBPTS, 3% for RHODE) Maximum: 9.7% (8.4% for NBPTS, 1.3% for RHODE	Improved district contribution to teacher retirement plan; early retirement option.
Douglas County	$300- 500 bonuses per skill block for 9 blocks; $1000 annual bonus for being designated an outstanding teacher.	Base: 17.4% Maximum: 6.9%	3% across the board pay increase; knowledge and skill part of plan voluntary
Limon	Entry pay based on a traditional seniority and credits schedule, but after entry progression based on an across the board increase, $1,000 for a Master's and $3,000 for a Ph.D., plus up to $1,200 in performance based increases, $400 of which is based on meeting individual professional development goals.	Base: 1.5% Maximum: 1.0%	Cost of living adjustment added to pay system, rectification of base pay inequities between new hires and more senior teachers.
Manitowoc	Expanded traditional salary schedule to provide more lanes and allow movement between lanes based on locally-developed courses and classroom-relevant university certification aligned to NBPTS, as well as traditional credits and degrees; 13% salary add-ons for NBPTS certification and Ph.D. degree. Seniority movement within a lane capped at lower pay levels to encourage obtaining advanced degrees and other recognized professional development.	Base: 13% NBPTS NA other parts Maximum: 13% NBPTS NA other parts	Improved funding of retiree health insurance premiums; new pay elements not covered by cost controls
Robbinsdale	Traditional salary schedule modified by reducing number of lanes from 11 to 4 and steps from 13 to 7. Knowledge and skill-based component provided for additional pay of up to $15,000, with the actual amount based on points earned in 10 categories: NBPTS certification, principal evaluation, individual accomplishments, district projects, contribution to teams, content knowledge, professional leadership, and customer satisfaction.	Base: 56.1% Maximum: 23.8%	To be applied to newly-hired teachers or volunteers
Vaughn	One 11-step seniority-based lane, $1,000 add-on for California teaching credential, $2,000 add-on for Masters degree, $2,000 add-on for qualifying as demonstrator for student teachers, $4,000 add-on for NBPTS certification. Three levels of competency-based pay add-ons (up to $13,100) earned by achieving a minimum score or better on rubrics in ten areas: Lesson planning and classroom management, literacy, language development, technology, special education inclusion, mathematics, history and social science, and science pedagogy, instruction in primary language for English learners, arts.	Base: 48.7% Maximum: 22.7%	Applied only to newly-hired teachers or volunteers from among veteran teachers

In all of the cases, some sort of quid quo pro or provision was added to sell the program, especially to highly senior teachers. It is interesting that in two of the cases, the consideration was relatively small. In Manitowoc, it was an additional district contribution toward retiree health insurance premiums. In Cincinnati, it was the exemption of a relatively small number of very senior teachers who would be likely to retire soon after the pay provisions took effect. In four cases (Robbinsdale, Vaughn, Douglas County, and Coventry) participation in the knowledge and skill-based pay part of the system was voluntary for all or senior teachers (though at Vaughn most of the senior teachers opted into the system in the second year). In Coventry, the district also increased its contribution to the teacher retirement plan and provided an early retirement option. In Limon, a small cost of living adjustment was added to pay system, and the school board corrected some base pay inequities that had emerged between new hires and more senior teachers. These experiences suggest that the potential opposition of senior teachers was an important issue to program designers. However, limiting the program to new teachers or volunteers may dilute the impact of the program on motivating improvements in instructional capacity. At Vaughn, the hope of the program's designers was that experienced teachers would volunteer to participate, and many did in the second year. In Robbinsdale, this was not a major concern because the primary impetus for designing the system was not to improve the skills of current teachers.

Both base pay increases and bonuses were used to reward knowledge and skill acquisition. Cincinnati, Coventry, Limon, Manitowoc, and Robbinsdale relied primarily on base pay increases, while the Vaughn and Douglas County knowledge and skill-based pay programs used bonuses. In some of the cases, however, some or all of the base pay increases were time limited "add-ons": the pay increase continued only for a fixed period, after which knowledge and skills had to be re-demonstrated. The programs that rewarded NBPTS certification provided the extra pay for the 10 year life of the certification. Increases based on locally- assessed knowledge and skills were provided for four to five years in three cases. At Vaughn, the base/bonus distinction was blurred because the bonuses are pro-rated and the extra pay is included as an add-on to the monthly base. This provided continuity of income for teachers, though the extra money needs to be re-earned every year.

Though the pay systems in these cases were diverse, one notable similarity across them was that six of seven retained a seniority-based element. (Limon eliminated seniority as a basis for progression after entry.) In this regard, most of these programs differed from private sector implementations of the knowledge and skill pay concept, which typically eliminate seniority as a basis for pay (Jenkins et al, 1992). However, in five of the cases where seniority increases remained, the new pay system decreased the emphasis on seniority by reducing the number of seniority steps or capping seniority-based pay progression at a lower level. Another similarity was that all of the systems continue to reward Master's degrees, suggesting that the K-12 sector continues to value higher educational attainment, and that moving away from this traditional valuation may be too radical a change to be accepted by teachers. It should be noted, however, that Cincinnati planned to limit pay increases for Master's degrees to those relevant to the teaching assignment.

The motivation model suggests that, all else equal, more substantial incentives will be more effective in motivating knowledge and skill acquisition. One way to assess the size of the incentive is to compare it to the entry-level salary rate and to the maximum salary pay a teacher can earn in a school or district. Table 5 contains estimates of the magnitude of the knowledge and skill incentive in the form of the percentage available for knowledge and skill-based elements (beyond those recognized in the traditional salary structure) as a percentage of the beginning base pay and as a percentage the highest pay rate available (including the knowledge and skill-based incentive, but exclusive of pay for additional activities like coaching). Again, significant variation existed, but it is clear that in three cases, Cincinnati, Robbinsdale, and Vaughn, the knowledge and skill incentive was substantial. The incentives provided by these three, as a percent of beginning pay, were on the order of those reported for private sector plans, which provide for 50 to 100% increases based on knowledge and skills (Gupta et al, 1986, Jenkins et al, 1992; Tucker and Cofsky, 1994.) The expectation that a larger incentive would be found where the motivation for implementation was to improve student achievement was partially fulfilled, in that some of the largest incentives were provided by Cincinnati and Vaughn. However, recruiting good teachers was the primary motivation in Robbinsdale, the other organization with a large incentive.

How the Acquisition of the Knowledge and Skills are Supported

Table 6 summarizes the professional development associated with the knowledge and skill-based pay programs, and the links between knowledge and skill-based pay and other aspects of the human resource management system.

Table 6
Integration with Professional Development and Other Human Resource Management Programs

Site	District Support for Acquiring the Knowledge & Skills Needed	Relationship to Other HR Programs
Cincinnati	While the district had an extensive professional development program covering many of the skills relevant to the teaching standards, there was no explicit linkage between the program and the standards that would allow teachers to determine which courses applied to each standard. District had new teacher mentoring and peer review programs that were being converted to use the teacher standards.	The knowledge and skill assessment system is same as used for teacher performance evaluation. At the time of our study there were no links to teacher recruitment and selection.
Coventry	Major changes to the professional development program were made to support improved instruction, but this was done before the knowledge and skill-based pay program was developed. Several courses have been developed to address procedural aspects of the knowledge and skill-based pay system, but otherwise there does not appear to be much explicit linkage between professional development and the pay program.	A modified version of the Framework for Teaching was used for teacher evaluation. This was also provided to job candidates as part of the recruitment process, and some interview questions are based on Framework elements. Though the district regarded the Framework as consistent with the NBPTS standards and the RHODE program, there was no formal link to the knowledge and skill-based pay programs.
Douglas County	Courses for skill blocks are provided by the district. Although the district offers a substantial number of other professional development classes, none are directly linked to the outstanding teacher program. Completing the portfolio itself was considered a form of professional development.	Originally, a connection with the state's multiple level licensing system was planned, but delays and changes in the state program prevented this development. Teacher evaluation was connected to regular pay progression, but the only explicit link to the knowledge and skill-based pay program was that teachers rated unsatisfactory cannot apply for the outstanding teacher award.

Limon	The overall professional development program was expanded at the same time the new pay program was developed. The nature of the individual professional growth goals left the choice of development activities to the teacher and supervisor, subject to general district guidelines.	Initial program tied pay to teacher evaluation system; current program no longer has the direct tie. The overall pay for performance system was explained to job candidates during recruitment; administrators felt this led to self-screening and higher retention.
Manitowoc	The knowledge and skill-based pay program was directly linked to a specified university professional development program and to locally-developed and provided courses. Existing local courses covered instruction, technology, and student writing.	No specific links between this program and other HR programs had yet been developed.
Robbinsdale	Due to delays in implementing the program, no specific skill-based pay program had been developed at the time of our study.	Due to delays in developing the system, connections with other HR systems were not been developed at the time of our study. Regular teacher evaluations were one element in teacher portfolio documenting knowledge and skill. Initially, the knowledge and skill- based pay program was used as a recruitment tool, but this ceased with the delay in implementation.
Vaughn	Some formal professional development provided around literacy and classroom management, provision of mentors for new teachers or those having difficulties meeting basic standards, and using coaching tied to the assessment process as part of skill building.	The knowledge and skill-based pay assessment standards were also used for teacher evaluation. The pay system was explained to job candidates as part of job interview. Administrators felt this was a recruiting advantage since most newer teachers could earn more at Vaughn than in the surrounding district.

At the time of our study, it did not appear that many of the programs had a strong professional development component specifically designed to provide teachers with the knowledge and skills rewarded. Those programs with larger incentives and more radical structures had not yet developed a corresponding comprehensive professional development programs. Vaughn had only begun to develop a comprehensive professional development program linked to the specific skills in the model. Formal professional development was provided on some domains, including literacy and classroom management, and mentoring and teaming were beginning to be used to help develop skills in the program domains. Cincinnati, though it had a comprehensive knowledge and skill model, had not yet modified their fairly extensive professional development program to link up with it. Robbinsdale's model was not yet fully fleshed out, but the diversity of the elements rewarded (ranging from principal evaluation to professional leadership and parent satisfaction) may make it difficult to identify specific skills and develop a coherent professional development program linked to the pay system. Those programs with tight links had smaller incentives and made less radical changes in the pay schedule. Manitowoc and Douglas County had the tightest links, in that parts of their pay program were directly tied to taking specific courses. However, both had not yet developed more than a relatively few courses covering a limited range of skills, and other professional development opportunities were not yet aligned with a comprehensive knowledge and skill model. The strategy of these two districts appeared to be to start small, paying first for attaining a few important skills. Coventry did not appear to link the professional development program to the pay program, perhaps due to the limited scope of the latter. The Limon program, and Douglas County's outstanding teacher award, gave teachers considerable choice as to what skills would be rewarded and therefore did not provide the basis for a comprehensive, linked professional development program.

At the point at which we studied these programs, the pay systems were not closely integrated with other human resource management activities. Two of the programs with the largest incentives, Cincinnati and Vaughn combined knowledge and skill assessment with teacher evaluation. Robbinsdale integrated the evaluation system by using it as one of eight elements in its assessment system. Vaughn used the knowledge and skill pay system in recruiting teachers, and Robbinsdale had planned to, but since the system was insufficiently developed, had not done so at the time of our study. None of the organizations appear to have used the knowledge and skill model in selecting teacher candidates at the time we studied them.

Additional Costs of the Programs and Methods of Funding

Table 7 shows the estimated additional costs, where available, of the knowledge and skill-based pay programs, and the method of funding these costs. It should be noted that the extra cost of salaries was hard to estimate, since there was little experience at most sites to tell how many teachers will move to the higher pay levels, and at what rate they will move. Therefore few solid costs estimates are shown. From the limited data provided, it appears that transition costs can be quite low, as can the costs in the first years before many teachers have had a chance to develop the full range of knowledge and skills. However, it is also clear that some of the programs provided the potential of substantially higher salaries. Comparing the maximum pay attainable under the former system with that attainable under the knowledge and skill-based pay system, a teacher in Cincinnati at the top of the schedule has the potential to achieve a 21% higher pay rate. In Manitowoc, such a teacher has the potential to earn 38% more, and Vaughn, 22% more. The other programs provided a substantially smaller additional pay opportunity. Limon provided only about 1.5% more pay, Coventry, about 11%, and Douglas County, about 7.4%. Robbinsdale's proposed plan provided for the same maximum as in the old schedule.

Table 7
Costs and Funding

Site	Costs of Pay and Administration	Source of Funds for Pay and Administration

Cincinnati	Transition cost to new pay schedule estimated at 0.2 - 0.4% of payroll; ultimate extra cost of pay changes not estimated. Cost of administration not known, but compensation for 8 full-time teachers to do classroom observations could be about $500,000 annually.	Re-allocation of some of the dollars spent on degrees and credits in the current pay schedule, reallocation of staff time and budget resources to administer the system. Some new money raised via higher local taxes.
Coventry	Estimate not available because program had just begun.	Most funding appeared to have come from increases in state funding. Reallocation of existing time and funds used to cover administration, most notably conversion of an administrator position to Director of Professional Development.
Douglas County	District estimate of cost of additional knowledge and skill pay elements was about 0.5% of payroll. No estimate of administrative costs is available.	Additional funds raised from local tax base.
Limon	District has not made an estimate, but if all teachers received the professional growth bonus, the cost would be about 1.4% of payroll.	Reallocation of existing funds and additional funds raised from local tax base.
Manitowoc	No estimate of additional salary costs solely due to knowledge and skill elements was available. Total package increase estimated at 1.5% to 2% of operating budget, and 3.8% of payroll. No additional administrative costs expected by district.	Local Academy was expected to be self-financing. New money available from tax base within legal limits used to finance pay costs.
Robbinsdale	No estimate available from district. Since the program would be applied initially to new teachers, immediate additional costs would likely be quite low.	Plan was to reallocate existing funds to cover additional pay costs.
Vaughn	Total performance plan cost about 3.5% of payroll in 1999-2000, expected to rise to 6% in 2000-2001. No estimate of administrative costs available, but some of the time of three new administrative positions should be considered part of the administrative cost.	Reallocation of savings from efficiencies in management and in managing funds provided by formula from the state and district; also, new money provided in the state funding formula was allocated to pay.

In none of these cases had estimates of additional administrative costs been made. The assumption appeared to be that the time and staff needed to administer the programs could be added to exiting workload or handled by reallocations of current staff. For several of the programs, this assumption will probably not prove problematic, because of limited scope or reliance on external assessments. For example, because it relied on discrete course grades and external certifications, the Manitowoc system requires little in the way of new administrative machinery, beyond record-keeping. The Limon, Douglas County, and Coventry systems required additional administrator time to review portfolios, but Douglas County, as described by Hall and Caffarella (1997), had not found this a major strain, perhaps because a relatively limited proportion of teachers participated in that part of the program. On the other hand, Cincinnati's experience, in a field test of the assessment system, suggested that most administrators did not have the time to do extensive evaluation (Milanowski and Kellor, 2000). In response, the district decided to hire eight teachers to specialize as evaluators. Vaughn, too, had initial difficulties finding the time for peers and administrators to do the required number of classroom observations. The school responded by reallocating positions to provide for more administrators, by increasing use of substitutes to free up the time of peer assessors, and by hiring two retired teachers as part-time assessors. As a charter school, Vaughn had a considerable amount of budgetary flexibility, and was able to tap grant funds to pay for part of these additional administrative costs. These experiences suggest that it is likely that knowledge and skill-based pay designs that use extensive internal assessment will require the allocation of additional resources to program administration.

Discussion

This article has attempted to summarize some of the main features of seven innovative teacher compensation programs that rewarded teachers for developing their knowledge and skills. Based on a simple theory of action, a model of motivation, and descriptions of private sector experience, a set of dimensions was developed to guide the analysis and comparison of the design of the programs. The major findings from the comparison are summarized below.

Motivation for change. There are a variety of reasons for designing knowledge and skill-based pay programs. Contrary to the assumptions underlying our theory of action, most programs were not primarily motivated by a desire to improve the knowledge and skills of the existing teacher workforce in order to improve student achievement, though some programs were motivated by a desire to recruit and retain more highly-skilled teachers and to support a particular vision of instruction. Other reasons illustrated in these cases were to respond to public pressure for a linkage of teacher pay to performance and to differentiate teacher pay based on teacher quality. State-level student assessment and accountability programs were an important factor in only those cases where student achievement was low. In most of the cases, however, student achievement was not considered a problem. Districts with high or acceptable student achievement appear just as likely to innovate, though the most comprehensive of the programs we studied were found where there was pressure to improve student achievement. Programs in Cincinnati and Coventry and at Vaughn seemed designed to motivate teachers to practice in certain ways, rather than to motivating them to develop specific skills. These programs more closely resemble private sector competency-based pay programs, which often include a more general performance component, while the skill blocks in Douglas County and the Manitowoc program resemble the skill-based pay model, in which the development of specified skills is rewarded.

Design process. In all cases, teachers participated in the design of the programs. The larger districts used formal committee structures outside of the collective bargaining process to involve relatively large numbers of teachers in developing the details. The smaller organizations were more likely to have designed their systems with less involvement. Contrary to expectations, some programs were designed within the negotiation process, though these tended to be the simpler ones. In all of our cases, a relatively high level of association-management cooperation, or trust between administrators and teachers, was present. Teacher compensation change is possible in a collective bargaining environment, and association or union leaders have been champions of the process.

Knowledge and skills rewarded. The knowledge and skills rewarded are generally those related to instruction, though none of the programs studied placed heavy emphasis on content-specific pedagogy. At the time we studied them, relatively few of the programs had defined an integrated model of the knowledge and skills needed for quality instruction, nor a progression of levels of skill development providing a path to mastery, though some of the organizations may have been moving incrementally toward such a model. There was no dominant method of knowledge and skill identification in these cases. Though most of the programs included the National Board standards as part of their model, the Board's standards were typically not highly integrated with the other knowledge and skills rewarded.

Knowledge and skill assessment. All of the programs use some form of performance assessment to assess the acquisition of at least some of the knowledge and skills rewarded, rather than relying completely on degrees or credits as indicators of teacher knowledge and skill. Five of the seven programs included external assessments, typically the NBPTS assessments. However, these assessments were not typically integrated into the assessment system as a check on internal assessments, or used as an indicator of a higher level of core teaching skills.

Size and structure of knowledge and skill incentives. As expected, there was some tendency for programs that were motivated by the need to improve student achievement to move farthest from the traditional schedule. These programs were likely to send the strongest motivational signals to teachers. However, in none of these cases were seniority and graduate degrees eliminated as a basis for pay progression. In four of the seven cases, movement away from the traditional salary concepts was incremental.

Support for knowledge and skill development. Few of the programs we studied have developed a coordinated professional development program that is specifically linked to the knowledge and skill model. Lack of alignment of professional development programs with the knowledge and skill model may reduce the motivational force of the rewards if teachers do not perceive they have the opportunities to acquire the knowledge and skills. That direct links to professional development programs are not strong may be due to the fact that the programs were not intended to remedy knowledge and skill deficits on the part of current staff. None of the programs have fully aligned their human resource management programs with a developmentally-sequenced knowledge and skill model. This suggests that the promise of alignment in fostering a shared conception of good teaching has not yet been fulfilled.

Costs and funding. The cost of transition to a knowledge and skill-based pay system appeared to be low in the short run, though costs are likely to increase over time to the point that new money will be needed to fund them. Little information on administrative costs was available, and in most of the cases, the increased administrative costs were met using existing resources. For the more ambitious programs, administrative costs are likely to be significant, and may not have been fully realized at the time of our study.

Implications for Research on Knowledge an Skill-Based Pay in the K- 12 Sector

Unlike the private sector, where skill-based pay and competency- based pay systems appear to have become relatively codified, there are multiple models of knowledge and skill-based pay in the K-12 sector. These various models were designed to serve a variety of purposes, not simply to support improved instruction. This implies that in evaluating the success of knowledge and skill-based pay programs, it will be important to take into account other program goals and to develop measures of program impact in addition to measures of instructional capacity or student achievement. For example, to the extent that recruitment and retention of highly-skilled teachers is an important goal, the quality of new hires and the degree to which more skilled teachers are retained and less skilled teachers leave will be an important outcome to measure. To the extent that programs are a response to community pressure for pay for performance or accountability, it may be necessary to look at community perceptions of the program. One rough indicator that Douglas County and Cincinnati informants mentioned was increased willingness on the part of the community to pass referenda providing more tax money for education. To the extent that the goal is to support the diffusion of a particular vision of teaching, measuring teacher acceptance and implementation of this vision will be important.

Of course, the most important outcome to many organizations considering developing and funding knowledge and skill-based pay programs is likely to be whether they are effective in motivating skill acquisition, changing instruction, and improving student achievement. But because knowledge and skill-based pay at this point encompasses such a variety of designs, it will be important to develop some measures of the potential causal "strength" of the program. A set of benchmarks could be developed, as was done by evaluators of the New American Schools implementation in Memphis (Smith et al, 1998, Ross, 2000). This would entail using the theory of action and motivational model to specify dimensions and develop some rubrics for judging how close the design and implementation comes to the ideal specified in the theory, then relating these ratings to measures of effects. For example, the theory of action and motivational model suggest that a program providing few professional development opportunities, little administrator and peer support for new skill acquisition, and relatively small incentives, would have a limited effect on instruction and student achievement. In such a case lack of evidence that knowledge and skill-based pay was associated with improved instruction or student achievement would not be surprising, but also would not provide much information about whether a stronger knowledge and skill-based pay design can help improve instruction.

It is interesting to speculate as to whether knowledge and skill- based pay in the K-12 sector will evolve toward a plurality of systems reflecting strategic district or school goals, local history, and designers' preferences, or toward a family of similar systems based on external, generic standards. In the private sector, the theoretical argument for knowledge and skill-based pay is that it can provide incentives to develop organization-specific skills that support a unique competitive strategy. The K-12 analog would be a set of schools operating under school choice or voucher systems, with little procedural regulation. Of our cases, the Vaughn charter school best fits this model and it did "tailor" its knowledge and skill model more closely to its mission, as set forth in its charter. But as argued above, in the K-12 sector the core skills are likely to be similar across schools and districts, and there is little competition across schools or districts. Thus over time we may see a tendency for convergence on external standards and assessments. The limiting factor appears to be a lack of external assessments aimed at differentiating among mid-career teachers.

Implications for Policy Makers and Program Designers

The experiences of the seven organizations we studied suggest a number of fairly clear lessons for the design of knowledge and skill- based pay programs. First, that even the most radical of the seven programs we studied retained seniority and degrees as pay criteria suggests that it may be unrealistic to expect completely performance- based pay systems to emerge. It may be necessary to retain some aspects of the traditional structure in order to have a realistic chance of implementing a pay system that rewards the acquisition of instruction-relevant knowledge and skills. As the Cincinnati and Vaughn cases illustrate, a program can be designed to provide significant incentives for knowledge and skill development while retaining some rewards for seniority.

Second, teachers' associations may be more open to changes in pay systems than administrators or school board members expect, but this openness is likely to be the product of high levels of trust developed through cooperation on other issues, and design features aimed at encouraging acceptance by senior teachers may be needed.

Third, it may be easier and faster to adapt a set of pre-existing teacher standards rather than to develop a knowledge and skill model from scratch.

Fourth, the transition costs to even a fairly extensive knowledge and skill-based pay structure can be low. It is, however, likely that administrative costs will be higher where an extensive system of internal skill assessment is used, and that payroll costs may significantly increase in the long term. So it is advisable to plan for ways to cover these costs.

It is also interesting to note that, for these pioneer organizations, state programs such as assessment and accountability systems or teacher licensing supported rather than drove teacher compensation change. Local issues, union-management relations, and the agendas of leaders were probably more important as initiators of change, and will likely be very important in sustaining and guiding a program until it has taken hold. For state-level policy makers who desire to use salary dollars more strategically to improve student achievement, one implication is that it maybe useful to provide a comprehensive and coherent model of knowledge and skills that are directly related to improving instruction for local organizations to adapt and customize. This would help focus teacher compensation change on strategically-important goals. The model could also be linked to state standards for students. If one way to improve student achievement toward state content standards is to ensure that teachers can develop and teach high quality standards-based curriculum units to all students (Cohen and Hill, 2000), then the model should emphasize the knowledge and skills needed to do this. Integration of the model with a multi- level licensing system and providing funds to increase pay for teachers with higher-level licenses could be another way for state-level policy makers to focus the system on strategically important goals.

Not only might state-level action help to encourage greater coherence and focus on improved instruction, but there may also be significant efficiencies to be gained from developing a state-level model rather than having each district or school work on the problem alone. External assessments could be developed for common knowledge and skill elements, to lower the burden on local schools and districts. One state that has been working along these lines is Iowa. In that state, education, business, and political leaders developed a comprehensive model for teacher performance evaluation, licensing, and compensation (Iowa Department of Education, 2000). Yet the ability of local organizations to customize a state model should be retained, in order to maximize the potential for local acceptance and to recognize that, at this early stage, no one knowledge and skill-based pay model has emerged as 'best practice'.

Acknowledgment

This article is primarily based on case studies done by staff researchers with the Consortium for Policy Research in Education's Teacher Compensation project at the University of Wisconsin-Madison. The original authors of the cases studies are: Eric Conti: Limon, Colorado, Robbinsdale, Minnesota, and Coventry, Rhode Island; Eric Conti, with the assistance of Eileen Kellor: Manitowoc, Wisconsin; Carolyn Kelley, Douglas County, Colorado; Eileen Kellor, Vaughn Next Century Learning Center, Los Angeles, California; Eileen Kellor and Allan Odden, Cincinnati, Ohio. I thank these researchers for allowing me to use their work, and to assure them and the reader that any errors in the descriptions, analysis, or conclusions are my responsibility.

The reader should be aware that staff of the Consortium for Policy Research in Education at University of Wisconsin-Madison were involved in the development of two of these programs. Alan Odden and Eileen Kellor provided consultation and technical assistance to the district committee that designed the Cincinnati program. Anthony Milanowski and Eileen Kellor have been involved in evaluations of the program paid for by the district. Allan Odden has provided technical assistance to staff at the Vaughn Next Century Learning Center, and Anthony Milanowski has provided occasional advice to Vaughn in the course of exploratory research on that school's program. Though the author does not believe that this involvement has biased his description or implicit evaluation of the programs discussed here, the reader may want to remember this involvement when reading this article.

Notes

The research reported here was supported by a grant from the U.S. Department of Education, Office of Educational Research and Improvement, National Institute on Educational Governance, Finance, Policy-Making and Management, to the Consortium for Policy Research in Education (CPRE) and the Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison (Grant No. OERI-R3086A60003). The opinions expressed are those of the authors and do not necessarily reflect the view of the National Institute on Educational Governance, Finance, Policy-Making and Management, Office of Educational Research and Improvement, U.S. Department of Education, the institutional partners of CPRE, or the Wisconsin Center for Education Research.

References

American Compensation Association (1996). *Raising the bar: Using competencies to enhance employee performance.* Scottsdale, AZ: Author.

Bembry, K. L., Jordan, H.R., Gomez, E., Anderson, M.C., & Mendro, R.L. (1998, April). *Policy implications of long-term teacher effects on student achievement.* Paper presented at the 1998 Annual Meeting of the American Educational Research Association, San Diego, CA.

Cohen, D. &, Hill, H. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record 102* (2), 294-343.

Conley, S.C., & Odden, A. (1995). Linking teacher compensation to teacher career development. *Educational Evaluation & Policy Analysis 17* (2), 219-237.

Daniels, D.R. (1997). Competency-based compensation approaches. In P. T. Chingos (Ed.), *Paying for performance: A guide to compensation management* (pp. 72-95). New York: Wiley.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Dwyer, C.A. (1998). Psychometrics of PRAXIS III: Classroom performance assessments. *Journal of Personnel Evaluation in Education 12* (2), 163-187.

Ferguson, R., & Ladd, H. (1996). How and why money matters: An analysis of Alabama schools. In H. Ladd, (Ed.), *Holding schools accountable* (pp. 265-298). Washington, DC: Brookings Institution.

Fisher, R., & Ury, W. (1981). *Getting to yes.* Boston: Houghton-Mifflin.

Gorsline, K. (1996). A competency profile for human resources: No more shoemaker's children. *Human Resource Management 35* (1), 53-66.

Gupta, N., Jenkins, G.D. Jr., & Curington, W.P. (1986). Paying for knowledge: Myths and realities. *National Productivity Review 5* (2), 107-123.

Hall, G. E., & Caffarella, E.P. (1997). *Third year implementation assessment of the Douglas County, Colorado school district performance pay plan for teachers.* No city: Authors.

Heneman R. L.. & Ledford, G.E. (1998). Competency pay for professionals and managers in business: A review and Implications for teachers. *Journal of Personnel Evaluation in Education 12* (2), 103-122.

Heneman, R. L., & Thomas, A.L. (1997). Using strategic performance management to drive brand leadership. *Compensation and Benefits Review 27* (6), 33-40.

Interstate New Teacher Assessment and Support Consortium. (1992). *Model standards for beginning teacher licensing and development: A resource for state dialog.* Washington, DC: Author.

Iowa Department of Education (2000, November). *A proposal for a new teacher compensation system based on teacher performance, career development, and the national labor market*. Des Moines, IA: Author.

Jaeger, R.J. (1998). Evaluating the psychometric qualities of the National Board for Professional Teaching Standards' Assessments: A methodological accounting. *Journal of Personnel Evaluation in Education 12* (2), 189-210.

Jenkins, G.D., Jr., Ledford, G.E., Jr., Gupta, N., & Doty, D.H. (1992). *Skill-based pay: Practices, payoffs, pitfalls and prescriptions*. Scottsdale, AZ: American Compensation Association.

Jones, T.W. (1995). Performance management in a changing context: Monsanto pioneers a competency-based developmental approach. *Human Resource Management 34* (3), 425-442.

Lawler, E.E. III (2000). *Rewarding excellence: Pay strategies for the new economy*. San Francisco: Jossey-Bass.

Ledford, G.E. Jr. (1991). The design of skill-based pay plans. In M. L. Rock & L.A. Berger (Eds.), *The compensation handbook* (3rd ed, pp. 199-217). New York: McGraw Hill.

Ledford. G.E., & Heneman, R.L. (2000). Pay for skills, knowledge, and competencies. In L. A. Berger & M. L. Rock (Eds.), *The compensation handbook* (4th ed.). New York: McGraw-Hill.

Mansfield, R.S. (1996). Building competency models: Approaches for HR professionals. *Human Resource Management 35* (1), 7-18.

Milanowski, A.T., & Kellor, E. (1999) *Vaughn Next Century Learning Center performance pay survey: Preliminary report*. Unpublished manuscript, University of Wisconsin- Madison.

Milanowski, A.T., & Kellor, E. (2000a). *Vaughn Next Century Learning Center performance pay survey: School report*. Unpublished manuscript, University of Wisconsin-Madison.

Milanowski, A.T., & Kellor, E. (2000b). *Teacher and evaluator reactions to standards-based teacher evaluation in the Cincinnati Public Schools: An evaluation of the 1999- 2000 field test of the new evaluation system*. Unpublished manuscript, University of Wisconsin-Madison.

Milanowski, A., Odden, A. & Youngs, P. (1998). Teacher knowledge and skill assessment and teacher compensation: An overview of measurement and linkage issues. *Journal of Personnel Evaluation in Education 12* (2), 83-101.

Mohrman, A., Mohrman, S.A., & Odden A. (1996). Aligning teacher compensation with systemic school reform: Skill-based pay and group -based performance rewards. *Educational Evaluation and Policy Analysis 18* (1), 51-71.

National Board for Professional Teaching Standards. (2001). State Incentives. Retrieved from http://www.nbpt.org/news/_center/facts/state_incentives.html.

National Board for Professional Teaching Standards. (1999). *What teachers should know and be able to do*. Arlington, VA: Author.

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. New York: Author.

Odden, A. (2000). New and better forms of teacher compensation are possible. *Phi Delta Kappan 81* (5), 361-366.

Odden, A., & Kelley, C. (1997). *Paying teachers for what they know and do: New and smarter compensation strategies to improve schools*. Thousand Oaks, CA: Corwin Press.

Ross, S.M. (2000). How to evaluate comprehensive school reform models. *Getting Better by Design Series 8*, 1-14. Arlington, VA: New American Schools Development Corporation.

Shippmann, J. S., Ash, R. A., Battista, M., Carr, L., Eyde, L. D., Hesketh, B., Kehoe, J., Pearlman, K., Prien, E. P., & Sanchez, J. I. (2000). The practice of competency modeling. *Personnel Psychology 53* (3), 703-740.

Slavin, R., & Fashola, O. (1998). *Show me the evidence! Proven and promising programs for America's schools*. Thousand Oaks, CA: Corwin Press.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review 57* (1), 1-22.

Smith, L., Ross, S., McNelis, M., Squires, M., Wasson, R., Maxwell, S., Weddle, K., Nath, L., Grehan, A., & Buggey, T. (1998). The Memphis restructuring initiative: Analysis of activities and outcomes that affect implementation success. *Education and Urban Society 30* (3), 296-325.

Spencer, L.M., & Spencer, S.M. (1993). *Competence at work: Models for superior performance*. New York: Wiley.

Tucker, S.A., & Cofsky, K.M. (1994). Competency- based pay on a banding platform. *ACA Journal 3* (1), 30-45.

Vroom, V.H. (1964). *Work and motivation*. New York: Wiley.

Wise, A.E., Darling-Hammond, L., McLaughlin, M.W., & Bernstein, H.T. (1984). *Teacher evaluation: A study of effective practices*. Santa Monica, CA: Rand Corporation.

Wilson, T.B., & Phalen, C.P. (1996). *Rewarding group performance: An approach to designing and implementing incentive pay programs*. Scottsdale, AZ: American Compensation Association,

Wright, S.P., Horn, S.P, & Sanders. W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education 11*, 57-67.

Zingheim, P.K., Ledford, G.E. Jr., & Schuster, J.R. (1996). Competencies and competency models: Does one size fit all? *ACA Journal 5* (1), 56-65.

About the Author

Anthony Milanowski
Associate Researcher
Consortium for Policy Research In Education
and
Industrial Relations Research Institute
University of Wisconsin-Madison
Madison, WI 53706
(608) 263-4260
amilanow@facstaff.wisc.edu

Anthony Milanowski is an Associate Researcher with Consortium for Policy Research in Education (CPRE) and a participating faculty member of the Industrial Relations Research Institute at the University of Wisconsin-Madison. He received his Ph.D. in Industrial Relations from the University of Wisconsin-Madison in 1997. Prior to joining CPRE, he worked in human resource management for 16 years, primarily with the Wisconsin Department of Employment Relations. His current research interests include teacher performance evaluation, pay system innovations, and the teacher labor market.

Appendix

Brief Descriptions of the Seven Knowledge and Skill-Based Pay (KSBP) Case Sites

Cincinnati

The Cincinnati Public School district is the State of Ohio's third largest, enrolling about 48000 students in 78 schools. The student population is 71% African-American, the rest white or other. About 65% are eligible for free or reduced price lunch. A relatively large proportion of Cincinnati's school-age children attend private schools, which, given the state's school funding system and laws requiring referenda for increases in school spending, has faced the district with pressures to reduce costs and improve student achievement. Average per- pupil spending was about $8,000 in 1998-99. The average teacher base salary was about $44,000. Fifty-one percent of the district's 3,000 teachers have Master's degrees and the average level of teaching experience is 15 years. The district has a seven-person elected school board on which the members serve staggered four year terms. Teachers are represented by a local American Federation of Teachers affiliate. CPS began developing its knowldge and skill-based pay program in 1996 with a commitment to redesign the teacher evaluation system. The new evaluation system, which will be the foundation for the KSBP program, was field-tested in the 1999-2000 school year and will be used in the 2001-2002 and 2002-2003 school years before the pay component. The pay component has been fully designed and was included in the collective bargaining agreement approved by the Board and the local Federation of Teachers in the Spring of 2000. In 2003-2004, the pay component will be added unless a supermajority of teachers vote to reject the program in May of 2003. In addition to the knowledge and skill-based pay structure, a group bonus of $1,400 is to be paid to all teachers in schools that meet school-wide goals for improving student achievement.

Coventry

The Coventry, Rhode Island district is one of the fastest growing suburban districts in the Northeast. It serves 5,600 students in 9 schools. The student population is 98% white and 2% minority. About 22 % are eligible for free or reduced price lunch. Student population has been growing moderately. Average per-pupil spending was $7,400 in 1997- 98, and the average teacher base salary was about $50,00. About 80% of the district's 351 teachers have Master's degrees and the distribution of teaching experience is bimodal, with about 60% very long service and 40% 5 or fewer years. Teachers are represented by a local American Federation of Teachers affiliate. Coventry began developing its new pay system in 1995, when the association proposed recognition of National Board certification. The initial element of the program, a bonus for NBPTS certification, was implemented in 1996. Additional pay for knowledge and skill elements were implemented for the 2000-2001 school year.

Douglas County

The Douglas County, Colorado school district, located in a fast- growing area between Denver and Colorado Springs, enrolls more than 32,000 students in 49 schools. (Thirty-two schools were opened since 1989.) The student population is 91% white, 4% Hispanic, and 5% other. About 2% are eligible for free or reduced price lunch. Average per- pupil spending was $7.817 in 1996-97, and the average teacher base salary about $39,680 in 1998. The average level of teaching experience is about 8 years. Teachers are represented by a local American Federation of Teachers affiliate. The process of developing the new pay system began in the 1991-92 school year, but the major design activities took place from July 1993 to the beginning of the 94-95 school year. The current plan was first implemented during the 1994-95 school year and has continued with minor modification since. Besides knowledge and skill elements, it also includes a modification of the traditional pay schedule that makes seniority pay progression dependent on satisfactory performance evaluation, a school group bonus program, and added pay for additional school-or district-level responsibilities.

Limon

The Limon, Colorado district serves 660 students in two schools. Located in a rural area, the district's students are 91% white, 5.6% Hispanic, and 3.4% other. About 34% are eligible for free or reduced price lunch. The size of the student population is now stable after an brief period of increase in the early 90's. Average per-pupil spending was $5,643 in 1996-97, and the average teacher base salary was $27,900 in 1998. 20% of the 44 teachers have Master's degrees, and the average level of teaching experience is about 10 years. Teachers are represented by a local association affiliated with the National Education Association, but only a minority of teachers pay state and local dues, and collective bargaining is essentially informal, with no formal contract negotiated. Pay innovation began in 1994-95, with the development of a link between pay increases and teacher performance evaluations. The traditional step and lane schedule was eliminated in favor of merit pay, which was based on principal evaluation in accordance with the state evaluation standards. The current plan was introduced for the 1998-99 school year. In addition to knowledge and skill-based pay, the pay system also the potential for a $400 increase if building-level goals are met and a $400 increase for achieving unit or grade-level goals.

Manitowoc

The Manitowoc, Wisconsin school district is located in a community of 33,000 in the eastern part of the state. It serves almost 6,000 students in 5 schools. The student population is 86% white and 14% minority. Student population growth has leveled off and is expected to decline. About 1% are eligible for free or reduced price lunch. Average per-pupil spending is about $7,692, and the average teacher base salary about $ 37,240. Relatively few of the district's 420 teachers have Master's degrees, but the average level of teaching experience is relatively high. Teachers are represented by a local National Education Association affiliate. Manitowoc began developing the new pay system in early 1999, as part of negotiations for the 1999-2001 teacher contract. The system went into effect for the 2000-2001 school year.

Robbinsdale

The Robbinsdale, Minnesota school district is located in a suburban area outside Minneapolis-Saint Paul. It serves approximately 14,000 students, of who 20% are nonwhite and 22 are eligible for free or reduced price lunch. (It should be noted, however, that within the district, the percentage eligible for free or reduced price lunch varies widely by school, from 15 to 80%.) The student population is slowly increasing. Average per-pupil spending is about $8,555, and the average teacher base salary about $ 44,950. 48% of the 900 teachers have Master's degrees and the average level of teaching experience is 14 years. Teachers are represented by a local American Federation of Teachers affiliate. The process of developing a new pay system began in 1994. Though the negotiation process, an outline of a plan was developed as part of the 1995-97 collective bargaining agreement, but the tentative contract was rejected by the membership. A revised program structure was approved as part of the 1997-2000 contract. The district is still working to develop the components, and the program has not yet gone into effect.

Vaughn

Vaughn Next Century Learning Center is a public charter school in San Fernando, California. Previously a public school in the Los Angeles Unified School Distinct, the school converted to charter status in July of 1993, It currently serves about 1,200 students in pre-K through grade 5. The student population is 94% Hispanic and 5.5% African American, and .5% other. Only 13% of the students are considered to be English proficient. About 98% are eligible for free or reduced-price lunch. The average teacher base salary was $42,000 in 1999-2000. The average level of teaching experience is 7 years. Vaughn is governed by three staff/parent committees, with a special council existing to oversee and resolve disputes between the three committees. Teachers as a group are not represented at this time, though some individual teachers are members of the American Federation of Teachers or National Education Association. The school began developing its KSBP program during the 1997-98 school year. An initial implementation for new teachers and volunteers was done in the 1998-99 school year. During that year, 19 of the 50+ teachers participated in the program. An improved version was put into place, again for new teachers and volunteers, for the 1999-2000 school year, during which 37 of the classroom teachers participated. The program was continued during the 2000-2001 school year. Vaughn's knowledge and skill-based pay is part of a complete redesign of pay system that also included pay for additional duties and a group bonus of $1,500 for all teachers if the school meets the student achievement goals in the charter. In addition, the school is eligible for a state program that provides bonuses to teachers in schools that meet state-set goals for improving student achievement.

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

This article has been retrieved 1539 times since February 6, 2003

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Education Policy Analysis Archives

Volume 11 Number 5 **February 6, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the **EDUCATION POLICY ANALYSIS ARCHIVES**. Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Continuing Education Reform in Hong Kong: Issues of Contextualization

Chris Dowson
Hong Kong Institute of Education

Peter Bodycott
Hong Kong Institute of Education

Allan Walker
The Chinese University of Hong Kong

David Coniam
The Chinese University of Hong Kong

Citation: Dowson, C., Bodycott, P., Walker, A., and Coniam, D. (2003, February 6.) Continuing education reform in Hong Kong: Issues of contextualization, *Education Policy Analysis Archives*, *11*(5). Retrieved [date] from http://epaa.asu.edu/epaa/v11n5/.

Abstract

Following initiations in educational reform that began in the 1990s, Hong Kong continues to experience considerable pressure for educational reform. On the surface many of these initiatives parallel reform policies/movements in Asia and indeed, globally. The success of any reform is dependent on how it is contextualised prior to and at implementation. In this article, an exploration is made into how reforms in four particular sareas, namely: professional development of principals, higher education, English language standards, and inclusion of students with learning difficulties have been conceived, contextualised and managed in Hong Kong, as it moves gradually toward increased adoption of education reforms. These areas are linked in that each describes and critiques contextualization with reference to areas such as accountability, co-operation and professional control.

Background to the Continuance of Education Reform in Hong Kong

Since the early to mid-1990s, Hong Kong, like many societies throughout the Asia-Pacific Region and beyond has been engaged in continual educational reform. The reforms cover almost the gamut of educational levels and issues. Reforms over the last decade have left few areas of education untouched. For example, they have included language teaching and learning, improving teacher quality, curriculum development, special education and various approaches to school-based management. The most recent reform initiatives are driven by the Blueprint for the 21st Century (Education Commission, 1999.), which sets the overall aims of education and maps a framework for reforms. The key elements in the Blueprint center around:

- expanding the opportunities for education and build a lifelong learning society
- introducing flexibility, diversity and choice in the system to accommodate individual differences and develop potential to its fullest
- creating an aspiring learning environment so that students are intrinsically motivated to explore and learn on their own
- introducing multi-dimensional assessments to encourage all-round development
- empowering frontline educators and enhancing the professionalism of principals and teachers.

Such efforts are focused directly on improving teaching and learning. The immediate tasks include the upgrading of language proficiency, enhancing school leadership and the professionalism of teachers, broadening access to education, and building flexible pathways for professional development. In short, recent and current reforms target increased decentralization, raising standards, increased accountability, equity and the building of professionalism. These efforts are just part of a very crowded educational reform environment in Hong Kong—one which influences educators at all levels.

The reforms introduced are, on the whole, typical of the educational reform environment in many countries in the Asia-Pacific region. Despite the vast social, cultural and political diversity of the Asia-Pacific Region there appear to be a cluster of reforms which seem remarkably similar, at least in espoused intent, across the region. These reforms say the right things and promise much. However, it is suggested that the success of any reform is dependent not on its rhetoric or its shape but how it is "negotiated" and implemented within a particular system or school.

In the following sections we will briefly explore how important reforms in Hong Kong have been conceived, contextualised and managed. The four reforms covered are:

1. The professional development of school principals
2. Higher education
3. English language standards
4. The inclusion of students with learning difficulties

Each section will describe and discuss (the process or content) of contextualisation with reference to areas such as accountability, co-operation and professional control.

Professional Development of Principals (PPD) in Hong Kong

Recognition of the key role of the school principal in education reform has grown substantially over the last decade or so. Such interest, at varying times and rates, is apparent in almost all societies in East and Southeast Asia. Educational reforms either targeting the role of the principal or areas that have an influence on this role are increasingly common. Accompanying such interest in the principalship are concerns that principals are unlikely to be able to play the role demanded of them unless they have the appropriate knowledge, skills, attributes and values required of reforming schools (Walker, Begley & Dimmock, 2000).

The reforms that continue to have the greatest influence on the role of the principal in the region are decentralization and the move toward school based management. Such reforms generally encapsulate school restructuring; school-based curriculum development; school development planning; increased teacher and parent involvement in decision-making and the formation of school councils. Other aspects may include delegated budgeting and human resource management; centralized curriculum planning using a learning outcomes framework; increased accountability to the central bureaucracy; increased parental choice of school; and greater competition between schools for students.

The gravity of these reforms has spawned considerable interest throughout the region in ways to help principals meet the emerging challenges of their role. For example, in Taiwan, Lin (2001) states, "reinventing schools requires exceptional school leaders—such leaders require a commensurate level of support and professional development to make the required role shift and, in many cases, this has not been forthcoming. This casts some doubt on whether they can adopt the new roles" (p. 8). In Japan, Muta (2000) comments, "The leadership and management skills of school principals are indispensable, but the current requirements for those positions are very strict, making it very difficult to find qualified persons" (p. 464). He adds "...some questions exist as to whether the principals can carry out such non-traditional tasks" (p. 464). In Hong Kong, Cheung (2000) expresses similar sentiments thus:

> Hong Kong principals face an uncertain, constantly changing and rather stressful future. Many are indeed over-worked as they face wave upon wave of reform initiatives. Additional responsibilities without adequate resources have made the role changes much more painful than necessary. How to find more room and time for principals to metamorphose into a new breed that can lead Hong Kong's schools triumphantly into the new millennium is an issue that needs pondering and concern by both authorities and the principals themselves. (p. 62).

Almost identical concerns in the UK and various states in the US and Australia have led to a proliferation of principal and leadership centers and subsequent increases in "training" opportunities for school principals. Regional societies have also publicly recognized that if schools are to change and improve, especially in an environment of decentralization and school-based management, then principals must be equipped to deal with uncertainty and actually lead, not just manage.

Policy makers in Hong Kong have recently moved to promote the further professional development of principals. Although the rhetoric of principal training and development has been present in Hong Kong for almost 10 years, little of any substance has actually been provided, except to small, targeted groups. On the whole, Principal Professional Development (PPD) has been a *loosely* structured and carried out on an *ad hoc* basis. In the main, its content and mode of operation has been decided and guided by the institutes of higher education; often based on their expertise rather than the needs of the principals themselves, or the reforms they are intended to implement. Alternatively, PPD courses have been prescribed by the Education Department and geared largely toward the technical/managerial skills considered necessary to run a school. However, recent moves by the Education Department (ED) have begun to lay a framework for a more holistic and coherent approach to principal development.

In January 1999 the Education Department established a task group to look into the training and development of school heads. This group developed a tentative program and framework "to equip and develop school principals with the necessary knowledge, skills and attributes to become competent leaders to lead schools into the new millennium" (Cheng, 2000, p.68). In the resultant consultation document Leadership Training Program for School Principals the group proposed the following objectives for program participants. These were to:

- assess personal leadership potential for further training and development
- increase understanding of the critical role of a principal in the development and maintenance of effective schools
- improve skills in strategic planning and implementation processes
- understand global developments and their implications for education and the school
- shape a personal vision for leadership and continuous development

While the consultation document was generally positively received, some interest groups expressed reservations about certain recommendations. These reservations included the difficulties instituting a "uniform" program for all principals (and potential principals) and the requirement that serving principals obtain a "certificate of principalship" by a set date. In reaction to these and other concerns, a second consultation document was released in February 2002. The consultation paper was entitled Continuing Professional Development for School Excellence (Education Department, 2002). This document presented a coherent framework for principals' professional development - one that aimed to meet the needs of Hong Kong practicing and aspiring principals at various stages of development.

Link to reform

The development and release of the two consultation documents and the promise of resources to support PPD by policy makers signaled recognition of the link between principal professional development and the successful implementation of educational reform in Hong Kong. It many ways, the consultation documents can be seen as the single most important move toward recognizing and systematizing principal training and development to come out for many years. As such, the initiative attempted to address a number of major concerns that have undermined previous efforts. Perhaps the most serious concern addressed by the emerging PPD policy was the lack of coherence and connectivity previously evident *between* and *within* the various components of PPD across levels. According to the document, PPD must target three distinct groups—Aspiring Principals (AP), Newly Appointed Principal (NAP) and Serving Principals (SP).

Linkage between the three levels is proposed through what the document refers to as the *Six Core Areas of Leadership*. These key areas of the principalship are common across the three levels of the principalship, although coverage and depth depends on the level. Attempts are also in train to build-in coherence between levels in terms of courses and other mechanisms. Linkage within levels is addressed through the inclusion of needs assessment, designated intensive and ongoing courses (partly based on outcomes of the assessment) and a certification processes. For example, according to the new policy, Newly Appointed Principals (NAP) begin their professional development by completing a needs assessment which is used, in turn, to inform group professional development offerings and individual development plans. Similar approaches are in-train or planned for APs and SPs.

Issues

Although there is no doubt that the current policy is a very positive forward step, a number of implementation issues hold the potential to impede the efficacy of the policy and accompanying programs. A first issue may be the number of different providers involved. Although this does not present a difficulty *per se*, it is obviously healthy to include multiple providers playing to their strengths, problems may arise if these providers disregard the established framework, or ignore the intended linkage between the needs assessment and designated programs, and design their inputs only in terms of their own interests. Such actions could weaken the effect of all the interrelated components and return principal professional development to its traditional fragmented and decontextualized roots. Another related articulation difficulty might arise between offerings and mechanisms for various groups of principals. Given that the framework divides principals into three groups, there appears a need to impose more stringent quality assurance mechanisms to prevent duplication and ensure relevance.

A further implementation blockage is that the current professional development initiatives are not linked to principal recruitment or selection. Aspiring principals and school governing bodies may not take PPD seriously if it remains de-linked from actually applying for and winning a position as principal. Unless principals and aspiring principals see a clearer, pragmatic purpose to professional development, such as selection or articulation to further degrees, the PD offerings may not be valued, or effective. Although it is hard to avoid, at the personal level there is also a danger of a lack of coherence, if insufficiently motivated individual principals engage in professional development in a very shallow manner.

How, why and where programs are conceptualized and developed may also hinder effective implementation. In many instances, there appears a continued, over-reliance on higher education institutions knowing what is best for aspiring and serving principals. Hence, although policy makers may reinforce relevance in terms of rhetoric, in reality, local providers (or individuals within certain organizations) may continue to develop professional development programs based on their expertise and ideologies rather than on what principals want and need to implement reforms. A result of this can be that offerings are overly formal (e.g. restricted to formal face-to-face courses and workshops), too practically biased (with little intellectual input) or, more often than not, overly theoretical and detached from the lives of principals. Such charges, unfortunately, are not uncommon in PPD programs where designers too often ignore the fact that much of what principals do is context specific, and that simple formulas, models or recipes distinctive to each context are not readily replicable. Until universities and other providers establish meaningful partnerships with principals at all stages of professional development there may well be continue to be a sizeable gap between offerings and the real life of schools.

A further blockage to the successful implementation of PPD in Hong Kong is the continued homage paid to overseas theories, presenters, frameworks and ideas, particularly English and American models (Walker & Dimmock, in press). As in many societies in the region, there is often an over-reliance on Western, mainly Anglo-American theory, values and beliefs in terms of overall policy adoption and in many areas of education. In terms of principal's professional development, unconsidered adoption can hamper the meaning of the program content for participants and influence its design, structure and even presentation. Although cross-fertilization of ideas and approaches is generally positive, there may not be enough recognition that theory, practice and imported expertise is often culturally insensitive or inappropriate. In short, it is difficult to see how PPD can be relevant if it depends almost exclusively on people and ideas located in very different contexts and cultures. One way to counter such a difficulty is to institute feedback mechanisms that provide local professionals (principals) with opportunities to comment, shift and adapt imported offerings to local needs. An even more effective mechanism is to ensure that programs respond to local needs as identified through systematic needs analysis.

Another implementation blockage relates to the political/micro-political environment within which professional development takes place. In many ways, this is related to the issues of coherence and continuity noted earlier. Historically, Hong Kong has practiced an established tradition of "sharing" out the provision of professional development to the various, relevant higher education institutions. Although unarticulated, this tradition has been based on a "keep everyone on-side and happy" philosophy. Such practices are antithetical to quality PPD and emphasize political rather than content/delivery concerns. Petty jealousies and competition between academic institutions can upset the intent and implementation of professional development. While concerns focus on matters other than the relevance and quality of professional development opportunities and provision, the efficacy of PPD may be questioned.

Finally, the implementation of PPD in Hong Kong may also be blocked by its lack of a "hard edge." In other words, the standard of accountability mechanisms may be inadequate both in terms of the principals involved in the process and for the providers. For example, there are no obvious mechanisms to gauge the effect of the professional development on leadership actually in the school. Neither are there stringent requirements for providers to show that their outcomes have in fact been achieved. Although the emerging policy is attempting to address this issue, particularly through the needs assessment process, it has some distance to go.

Recent policy then, is aimed at the professional development of aspiring, new and serving principals in Hong Kong holds previously unequaled opportunities to build principal professionalism. The scope of the changes calls for a marked shift in culture in terms of how principals, providers and the system perceive and operationally PPD. At it's most basic level, the new policy requires a change in mind-set—away from the belief that principals do not need to learn, or learn by osmosis. For principals to successfully reshape their role they must have access to meaningful professional development. For such development to become relevant and offer any chance of real change it should be developed in concert with principals themselves and be adequately resourced and rewarded by departments and ministries of education. It should also be linked closely to the reforms principals are expected to implement and shaped to form a coherent program rather than the piecemeal, fragmented attempts that comprise the norm.

Much of the responsibility for making professional development meaningful, of course, lies with principals themselves and their willingness to take some control of their own destinies. Although the emerging PPD policy holds much promise, it faces a number of fairly major hurdles if it is to be implemented in a way that truly benefits principals and cajoles them toward greater professionalism and reform.

Teacher Education Reform: The Challenge of Contextualisation

Over the past 25 years there has been an explosion of knowledge about brain development, cognition, learning theory and their relationship to student motivation and achievement (Lynch, 2000). In Hong Kong this knowledge has gradually made its way into educational reform documentation that posed and continues to pose significant challenges to the way the curriculum is structured and the ways children are taught. The challenge for teacher educators is to figure out how to incorporate new knowledge, required by reform initiatives, in to our teacher education curriculum, in order that new teachers understand how ultimately to advance student achievement in their classrooms. This is a difficult task, made more difficult in Hong Kong by reform initiatives that apparently contradict one another, and by a school system that does not necessarily welcome students well versed and skilled in new approaches to learning, teaching and assessment, or equipped with a notion of professionalism, which includes the right to question existing practice.

In this section of the paper it will be argued that a paradox exists between the requirements of reform and the needs of schools. This paradox places, for one institution specifically, undue political demands and expectations on the teacher education curriculum, teacher educators and student teachers. It is also argued that to overcome the apparent paradox, teacher education curriculum reform is required in which issues of contextualisation are addressed.

The Reform Context

Recently released education reforms (Education Commission (EC) 2000) comprise wide scale changes to all sectors of education in Hong Kong. The reforms encompass a broad vision based on the principles of learner-centered education and related changes to the curriculum, modes and focus of instruction, assessment opportunities, whole-school management, enhanced professionalism, and opportunities for lifelong learning. The reforms also embrace the worldwide trend toward accountability through standard-based assessment. Inherent in the reforms are recommended changes or modifications to all stages of education: Early Childhood to Higher Education and Continuing Education. More substantial than the mere scope of reform is the depth of reform, which strikes at the ideological heart of education in Hong Kong. The reforms are summarized in the following five principles.

- Learning should be focused on students' personal development and allowing room and flexibility for students to be masters of their own learning.
- There should be no "dead-end screening" to block learning opportunities. Learners should be given due recognition for what they achieve.
- Access should be provided and means sought to ensure learners realize their potentials. "Everybody should achieve basic standards and strive for excellence."
- Learning should not be limited to school subjects or examination syllabuses, students should be prepared for the realities of life
- All sectors of the community are expected to contribute to the reforms.
  (EC, 2000)

It is clear that the reforms aim at improving teaching learning standards and accountability across all stages of education through changes and improvements in academic structure, curriculum, and instruction and assessment mechanisms at the various levels. However, while several of the reforms, such as the development of technological expertise and meeting the needs of diverse learners appear compatible with general developments in higher education and school curriculum, other aspects appear to contradict one another, e.g. learner-centered teaching based on a premise of teaching for understanding and standards-based assessment. In addition, there are problems associated with competing ideologies and related perceptions of educational aims and practices between policy makers, and policy implementers, (which are the schools). These apparent contradictions and differences of opinion raise serious questions regarding contextualisation.

The reform documentation clearly acknowledges a desire to shift classroom instruction away from the transmission of knowledge to learning how to learn i.e., establishing a new culture of learning and teaching. However, questions must be raised about contextual or cultural relevance of such a movement. For modern Chinese, Confucianism remains a critical element in their cultural identity (Chan, 1999). Closely linked to this is a traditional view of the teacher as authoritarian and one whose role is that of transmitting knowledge through instructional methods such as rote and repetitive learning. Such methods are reinforced by an emphasis by the teacher and parents on demonstrable knowledge achievements in the form of test score attainment. Therefore, to propose educational reforms that clearly contrast with firmly established cultural styles of learning and teaching, through assertions such as "the fostering of learning abilities [to be] more important than the imparting of knowledge" creates something of a educational paradox; to marry the needs of continuous change through reform and innovation on one hand to an educational system that fundamentally rewards conformity and orthodoxy on the other. In Hong Kong, politics also plays a significant role in deepening the paradoxical gap between preparing reform-savvy beginning teachers and the contextual reality of schools.

Contextualisation in respect to teacher education and general education reform is defined as the negotiation and adaptation of a policy, programme or curriculum approach for a more meaningful fit to the values, norms and structures of a host culture or organization (Bodycott, in press). The concept and more importantly the issues arising from its application pose one the greatest challenges for any teacher education institution intent on creating and maintaining change.

Contextualisation and Teacher Education in Hong Kong

Higher education in Hong Kong refers to all learning opportunities above secondary school level. This includes post-secondary school colleges, universities, extra-mural departments of universities, non-local tertiary institutions and the Hong Kong Institute of Education (HKIEd), a teacher education facility and itself a successful product of EC Report 5 in 1992, and the main source of primary school teachers for the HKSAR.

The challenge and associated problems posed to teacher education institutions by the reforms in respect to curriculum; instruction and assessment are similar to those discussed generally above. That is, not to confine curriculum to the transmission of knowledge and skills, but to provide students with training in aspects of culture, emotion, moral conscience and mentality. Specifically, the reform recommendations require all higher education institutions to "review the functions, contents, focuses and modes of teaching… to strike a balance between the [content] breadth and depth" (EC, 2000, p. 113). Higher education must provide learning experiences and opportunities to "develop-broad based knowledge and vision, as well as enhancement of individual students problem- solving power and adaptability" (EC, 2000, p. 111). Hong Kong institutions of higher education are being challenged to rethink their whole approach to curriculum and instruction.

Pedagogical issues

The reforms also constitute for teacher educators theoretical and pedagogical dilemmas based on the incongruent principles of reform and their explicit emphasis on student-centered approaches to teaching and learning on the one hand, and the sustained emphasis on standards and outcomes assessment on the other. The problem becomes professionally challenging when dealing with an overt clash of philosophical ideals—standards-based accountability and liberal-humanist values and the translation of these into programmes of instruction that meet the contextualised needs of schools and general education (Helsby, 1999). The "imposition" of standards- based accountability locks teachers into a mode of operating in schools, no matter what efforts are made to reform or shape it (Apple, 2001). In colloquial terms, teacher educators in Hong Kong are metaphorically, caught between a rock and a hard place. The reforms charge teacher educators with the responsibility of creating "pillars of society" who are both "generalists and specialists" who will have high standards of academic knowledge and language skills, and who (as beginning teachers) are entering a school system in a time of rapid reform.

Politics, Reform and Paradox

In Hong Kong an undercurrent of concern exists relating to the politicization of the reforms and specifically their direct effect on teacher education. If anyone was to question the importance placed on the teacher education in respect to the latest reforms, then they need look no further than to the level and tone of language used to describe their role. Higher education is seen to be "the key" and has "the duty" to be the gatekeepers of among other things student language proficiency attainment. The reform documentation describes the move toward exit standards of language proficiency. Such language and initiatives firmly position the current reforms alongside similar global trends aimed at national standards for teachers e.g., in the United States and in Australia. But in these countries such movements are based largely on economic trends and questionable motives and place extreme pressure on providers of teacher education. Whatever the political motive, the reforms and in particular the emphasis on outcomes-based performance assessment and the associated high stakes tests for graduating language teachers in Hong Kong has already exerted a considerable influence, albeit not always positive, on teachers and teacher education students, and higher education teaching, curriculum and teacher education agendas.

Currently in Hong Kong the majority of newly trained primary teachers are provided by the HKIEd. Two major universities compete with the HKIEd for this substantive market. However, an inequity exists between the universities and the HKIEd, which is compounded by the paradoxical relationship between reforms, teacher education and schools. Specifically, issues arise concerning the role and status of the HKIEd and its ability to prepare teachers for the contextual reality of Hong Kong schools, to "compete" for student numbers and the necessity for it to demonstrate programme conformity to government reform directives. These issues, as summarized in Bodycott, (in press) include:

The status afforded HKIEd where in an elitist educational context is yet to be awarded "university" status, which affects the level of clientele attracted to programmes of teacher education at HKIEd compared to other "university" education faculty programmes. HKIEd and teaching tends to attract students who are generally of lower academic standard than those entering for study at "universities." Therefore the impact of the unstated competencies and "university standards" expected of all graduating HKIEd students is far greater given their entry level.

Secondly, the requirement of the HKIEd to have full details of all programme content and management validated for rigor and general quality by external parties including international subject and curriculum experts and local academics from "university" faculties. These validations require the HKIEd to defend all aspects of design and management and in particular how the programme reflects government reform initiatives. No other teacher education faculty in Hong Kong is required to divulge such programme detail or justify relevance and relationship of specific teaching philosophy and subject content to government reform directions.

Thirdly the perceived need of government and the community to criticize education in general and the HKIEd specifically for failing to deliver students and beginning teachers who are academically, socially, spiritually and morally fulfilling of their, at times, unrealistic expectations.

Lastly the reforms create inequities in the provision of teacher education. University faculties of education via their status exercise liberties in respect to programme design and the degree in which government reform initiatives are addressed, whereas HKIEd programmes by virtue of it's "non-university status" must conform to government reforms. The paradox here is that schools criticize the graduates and their programmes of instruction for not being compliant with the current pre-reform specific needs of schools.

Transforming Teacher Education

Current reform requirements and expectations imposed on higher education in Hong Kong presents teacher educators and their students with what many find almost insurmountable challenges. Firstly, to meet the contextualised demands of reform teacher educators must themselves be capable of personally and professionally dealing with change and with the inherent ideological and pedagogical incongruence associated with reform. Secondly, assuming teacher educators can personally cope with such challenges they then must have or develop the professional ability to transpose their understandings into programmes of instruction that can in some way challenge the existing beliefs, attitudes, values and practices of students. To meet such personal and professional challenges it is proposed that there must be a significant change in the nature of teacher education. At present, missing from the reform agendas in Hong Kong and elsewhere is a view of teacher education that encompasses a "change agentry" (Fullan 1993). That is, time must be given within teacher education programmes to overtly prepare teachers and teacher educators to be effective agents of change.

Preparing teachers to be agents of change is decidedly not the focus of current teacher education in Hong Kong, nor importantly is it necessarily desired by future employers. Principals, despite educational reform initiatives, do not necessarily want new teachers to enter their schools questioning existing practices, or reflecting and acting responsively to the culturally relevant academic needs of differentiated learners and by doing so upsetting the status quo. Frequent criticisms by Hong Kong principals of beginning teachers, and especially those from the HKIEd, is that they are under-prepared, unprofessional and lack classroom management schools (criticism based on the noise- levels generated in classrooms), and lack the subject- knowledge to adequately prepare students for high stake tests.

The contextual reality is that generally Hong Kong principals do not want teachers to question whose interests are being served and met by school policy, or by the curriculum and its reliance on textbook teaching and test preparation. Nor do they want beginning teachers who may question the failure of the school to adhere to principles such as whole person, no loser, broad-based knowledge and learner-centered approach to education, as presented in the current education reform documents. Such knowledge of schools and the contextual challenges of teaching in Hong Kong are bound to have far reaching effects on the long term transformation of curriculum or practice and the design and effectiveness of any teacher education programme.

Placing Context to the Fore in Teacher Education

To have face validity, teacher education programmes must be contextually authentic. That is, students must be required to undertake a programme of study and teaching experiences that relate directly and authentically to the context of schools and the classroom. Such a belief is grounded in experience and the literature (for example see Lynch, 2000) which advocates the when students are taught in active, engaging environments, and are allowed to put the knowledge in to practice by demonstrating its application in some way, then they will come to understand more, retain more, and apply their knowledge more in various learning contexts. Such instruction Howey (1998) labels as "contextual teaching and learning (CT&L)" which is defined as:

> Contextual teaching …enables learning in which students employ their academic understandings and abilities in a variety of in- and out-of-school contexts to solve simulated or real-world problems, both alone and with others. Activities in which teachers use contextual teaching strategies help students make connections with their roles and responsibilities as family members, citizens, students, and workers. Learning through and in these kinds of activities is commonly characterized as problem based, self-regulated, occurring in a variety of contexts including the community and work sites, involving teams or learning groups, and responsive to a host of diverse learners needs and interests. Further contextual teaching and learning emphasize higher-level thinking, knowledge transfer, and the collection, analysis, and synthesis of information from multiple sources and viewpoints. CT&L includes authentic assessment, which is derived from multiple sources, ongoing, and blended with instruction. (pp. 19-20)

A contextually authentic teacher education programme would consist of the following features:

- The standards for excellence in teacher education would be clearly articulated and addressed specifically to students throughout their programme.
- Emphasis would be placed on developing the individual skills and qualities of students through an acknowledgement that reform or change is a highly personal process with no guaranteed outcome.
- Grounded in a personal-social-constructivist view of knowledge and skill formation, reform initiatives are channeled through contextualised tasks and activities designed to facilitate challenge to each individual students personal and professional constructs of teaching and learning. Through such a process, knowledge (theory), values and practices are analyzed and critiqued for what they actually mean for the student personally and in respect to the contextualised realities of classroom instruction, teaching and learning and other aspects of teachers work—policy formation, curriculum development, assessment and reporting.
- Therefore the development of reflection would be more than a rhetorical component in the programme design.
- A focus on assisting students to develop an understanding of the synergistic nature of theories in practice, of professional partnerships with fellow stakeholders in the education endeavor, of reflexive thought and action and the often incongruence of educational vision and contextual reality.
- All teachers would be expected to operate as reflective practitioners and thereby work collaboratively in learning communities and demonstrate that their teaching leads to increased student achievement (Cochran-Smith, 2001).
- Students would ultimately be expected to present themselves publicly, to justify the decisions they make in respect to teaching and learning to various stakeholders in the education process. These stakeholders include

parents, teacher colleagues, the principalship, school benefactors and the children themselves;.

- Authentic assessments are implemented that require students to actually perform certain learning tasks, thereby demonstrating their skill and understandings and an ability to apply what they have learned (McTighe, 1997). Such assessments require students to perform the assessment in a manner in which stakeholders wish them to perform in the classroom. In so doing, students are required to focus on higher levels of cognitive complexity (Gipps, 1995).
- Success would be determined through an evaluation of how well the programme content, assessment tasks and associated experiences reflect the real-world of the classroom.

Contextual authenticity in programme design and delivery would provide students with the best possible teacher preparation and the education community with teachers who are well-versed and skilled in the needs of the profession and the schools and children in the community and culture in which they are to serve. In a contextually authentic programme, students are placed in situations that require the application of knowledge and skills together with guided reflection and constructive feedback, as opposed to more traditional teacher preparation programmes that place greater emphasis on the recognition of subject content and reproduction of correct answers.

The roots of contextually authentic teacher education are grounded in theoretical principles of humanism and constructivism, which ensures students are "in touch with their own landscape" (Greene, 1978, p. 39). Contextually authentic teacher education embraces other theories and terms used by learning theorists—such as experiential learning, real-world learning, active learning, learner-centered instruction, and action learning—all of which are frequently quoted in reform documents if not observed in classroom practices.

The extent to which such principles influence the design of teacher education programmes will vary according to culture and education context. For example, if the agreed focus of education is mastery of techniques or demonstration of specific competencies, then teacher education programmes will need to be designed to transmit the required knowledge and in so doing prepare their prospective teachers in accordance. Similarly, if the focus or aim of education is for teachers to be more critical, reflective thinkers and practitioners—as in the current Hong Kong educational reforms—then prospective teachers must have opportunities to learn how to explore the world of teaching and education from a reflectively critical perspective. In so doing, they would be required to find their place within such a landscape before expecting their students to do so.

To be contextually relevant, teacher education must prepare prospective teachers to negotiate the blurred and somewhat contradictory realities of curriculum policy and school-based practices. Students must be prepared to prove themselves competent in their first time teaching positions while at the same time be prepared to challenge some of the assumptions and actions that other teachers and principals take for granted. It is believed that contextually authentic teacher education programmes would influence the students and schools affectively. When students and the school community perceive programmes as having personal and real-world "contextual" relevance they are more likely to feel positive about the programme and thus put more effort into it, and consequently principals would be more willing to employ teachers who graduate from it.

Unfortunately the most recent reforms in Hong Kong like many before them, may well be ignored, or even lost if apparent ideological contradictions are not addressed. Simply, beginning teachers confused by the competing ideological expectations and rewarded by outcomes-based incentives in their schools may simply choose to maintain the status quo and teach the way they have always taught (Bodycott, in press). While it is relatively easy for governments to formulate and initiate reform policies, the true test of their applicability comes at the implementation level (Walker, in press). The contradictory nature of the current educational reforms in Hong Kong place higher education in general and teacher educators in particular in an invidious situation. For in an era of high contextual expectation they are charged with preparing new and in-service teachers for schools, where curriculum and instructional visions are not aligned with cultural and contextual expectations regarding teaching, assessment and accountability. With government unlikely to change their educational reform focus, despite the irregularities, the answer for Hong Kong at least, lies in part with the development of a more contextualised approach to teacher education.

English language teacher standards

Standards of education in Hong Kong have been a cause of concern since education's early days (Bickley, 1997). It was really in the late 1980s, however, that the language standards of teachers in Hong Kong, particularly teachers of English, became an issue of general concern. The business community (Au, 1998; see also Choi, 1998), in particular, felt that English language standards were dropping among the workforce. This was worrying because higher standards were required as commerce moved from a predominantly light-manufacturing base to a service-led economy, one that dealt with the world on a regular daily basis through the medium of English. (It should be noted that the establishment of teacher language benchmarks has not been not the only measure to improve standards of teaching and learning: there have been a number of commendable initiatives designed to improve the curriculum and examination systems in Hong Kong—Falvey and Coniam, 2000).

Mindful of concerns about standards, the Hong Kong Education Commission (1), in December 1995, published Education Commission Report Number 6 (ECR6). The Education Commission highlighted a number of areas for action in this report, one of which concerned teacher competencies, in particular the upgrading pf *teacher* language standards. The establishment of language "benchmarks" (i.e., minimum standards of ability in language) was recommended for all teachers in Hong Kong (there are approximately 42,000 primary and secondary school teachers, of which approximately 12,500 are English language teachers) on two fronts. The first concerned language teachers, that is, teachers of English, Chinese and Putonghua. The second concerned teachers who teach content subjects (history, geography, biology, mathematics etc.) through the mediums of either English or Chinese (see Falvey and Coniam, 1997).

As a follow-up to the recommendations in ECR6, an investigative consultancy study (Coniam and Falvey, 1996) was commissioned in early 1996 to investigate the feasibility of establishing language benchmarks, initially for lower secondary (i.e., Grades 7-9) teachers of English (2). Following on from the consultancy report, steps were then taken to develop the recommendations of the consultancy report into government policy. The first, and most important step in developing policy, involved the creation, in late 1997, of a widely representative English Language Benchmark Subject Committee (ELBSC). The ELBSC was composed of the main stakeholders involved in English language education in Hong Kong—language teacher educators, members of the Hong Kong Education Department, the Hong Kong Examinations Authority, school principals, department heads, practising teachers and members from ACTEQ itself. The body, which was excluded from the ELBSC, was the Teachers' Unions—a move that had repercussions later when the benchmark policy was formalized and publicized.

Process

The ELBSC worked together or in sub-committees over the period 1997-2000 agreeing on assessment constructs, establishing specifications, designing exemplar tasks, creating scales and descriptors for criterion-referenced task assessment and monitoring the piloting and moderation of the assessment instruments. The work of the ELBSC culminated in the recommendation of a benchmark examination consisting of battery of "formal" tests (i.e., Reading, Writing, Listening and Speaking), as well as a performance test of Classroom Language, where teachers would be assessed teaching two of their own classes in their own school.

Following the work of the ELBSC, the HKSAR Government publicized the introduction of language benchmarks in mid-2000. The policy document stated that pre-service teachers would, from September 2001, have to be benchmarked before joining the teaching profession. In-service teachers, i.e., established serving teachers, would have until 2005 to meet the prescribed benchmarks. It is important to note, however, that the Government's drive to set a minimum standard of language proficiency did not reside solely in a test of teachers' language ability. In addition to the administration of the benchmark assessment, substantial financial resources were allocated so that every teacher of English could enroll on a language enhancement programme of up to a maximum of 200 hours (3).

It is worthwhile briefly examining the benchmark performance test—the Classroom Language Assessment (CLA) test, due to its innovative nature, the demands it places upon teachers and their expectations of the test. The CLA was discussed at length in the ELBSC because it would be a performance-based test that would take place in a live taught class. As Sanaoui (1999) notes, the essence of attempting to define what is fair, yet what also needs to be assessed in a performance test such as the CLA, lies in the form and manner of the assessment being determined "by consensus" across a group of informed stakeholders. The composition of the ELBSC, with teachers, principals, ED members and tertiary institution lecturers was an attempt to reach an "informed consensus." While the ELBSC was very much in agreement with the philosophy behind using an authentic test, logistic concerns were expressed at the administration of a live CLA. Although English language teachers are used to paper-and-pencil tests (preparing their students for such tests in public examinations), and formal tests are an accepted part of school culture, a live classroom test would be much more threatening. On this basis, such a test had to be carefully handled: the constructs assessed had to be broad in terms of the language skills assessed, i.e., not biased against any particular group—primary versus secondary, for example. Support for the retention of CLA was made in a 1999 Colloquium on English Language Benchmarks held in Hong Kong, where Nevo, a visiting scholar in assessment and evaluation from Israel, stated unequivocally that the inclusion of the CLA in language benchmarking should be retained in spite of inevitable arguments that it would be costly and time-consuming. His assertion rested on the proposition that CLA is at the heart of the teacher language performance being assessed (for a discussion of the CLA test, see Coniam and Falvey, 1999).

In March 2001, the first, live benchmark test for English language teachers was administered. When the results were released in June 2001, there was an outcry in the local media because of the apparently low pass rates: the headline of the South China Morning Post of June 9, 2001 stated *Teachers flunk English test*. The lowest pass rate was for the Writing Test, which 33.3% of test takers had passed.

Critique : Principles and problems

A brief examination of the benchmark initiative from the perspective of its positive and negative aspects follows.

On the positive side it should be noted that the benchmark initiative is a step forward in the area of teacher professionalism. The setting of language standards is a prerequisite for able language teaching. While it is understood that good language skills themselves cannot be necessarily be equated directly with "good" teaching, if communicative language teaching—whereby ESL teachers use a considerable amount of the target language to interact with their students—then a high degree of competence in the second language is essential.

In order that greater professionalism is viewed from a perspective of *encouragement* rather than simply as a stick waved by government, without resources being available, the move becomes one, which is even more, resented. The benchmark initiative has not of course been welcomed by all teachers—some, the less able it is often suggested feel threatened by the initiative. Nonetheless, the fact that the government has allocated financial resources (a total of US$30) for, in principle, every single teacher in the SAR is an indication of the importance the government attaches to the initiative in terms of the professional upgrading of teachers.

The test specifications, although initiated by Government were set by the consensus of a large and generally representative committee, with the omission, as mentioned above, of the teachers' unions.

On the negative side, as might be expected with a large- scale government initiative, the time frame has been, to say the least, unrealistic. From the initial moves with the 1996 consultancy study, the first administration of the first examination took place in 2001. This five-year lead- in contrasts with major syllabus revision in the case of public examinations, where in Hong Kong at least, a 7-year lead-in is the norm.

The insufficient time frame has also meant that all test types have not been trailed as extensively as they might have been—resulting in outcry from teachers and teachers' unions after the first administration in 2001. The exclusion of the largest teachers' union, the Professional Teachers' Union (PTU), resulted in initial work being completed without major agreement. That said, the hurdles were then even more difficult to surmount because of the earlier exclusion of the PTU. Sanaoui (1999) notes that for acceptance of an initiative to be embraced, the investigating body must comprise all stakeholders. The fact that pass rates were low in the first live administration of the test may in part be ascribed to the lack of readiness for the test among test takers. In order to pass the benchmark test, however, the ELBSC recommended that a pass must be achieved on every subtest. While this, in essence, was a laudable recommendation, in that it was intended to help "raise standards," the initial low pass rates which were very poorly received by the media did not raise much public sympathy for teachers. Further, it is argued that such a stipulation (i.e., having to pass every subtest) does in fact require a "minimum acceptable standard"—the original purpose of the benchmark initiative—but demands a higher than "minimum" standard (see Coniam and Falvey 2001 for a discussion).

Further, goalposts changed as the initiative developed: at first it was decided there would be no exemptions from the test. The publication and public promulgation of the test specifications in mid 2000 resulted in a lot of opposition. In an attempt to pacify the some of this opposition, an investigative panel was set up, which did then recommend exemptions.

A final point concerns the syllabus specifications not transparent enough—with performance tests such as CLA, print examples are not sufficiently transparent or informative. The same principles of accessibility can be found in the syllabus document produced for the English language benchmark assessment initiative. In a print document—such as that produced for the English language benchmark—specifying the exact demands of a performance test such as the CLA component is not easily accomplished. The Hong Kong Examinations Authority and Education Department have gone to great lengths to produce a detailed syllabus document. The document contains the scales and descriptors against which test takers will be assessed and contains as many exemplars of the Reading Test, Listening Test and Writing Test as was possible to put together in the limited time frame prior to publication of the benchmark tests. See Coniam (2002) for a discussion of the problems associated with a print syllabus attempting to specify the demands and requirements of an oral performance test. And an attempt to render the CLA test more accessible though a multimedia implementation.

Inclusion for Students with Learning Difficulties

Introduction

Current educational reform in Hong Kong includes a consideration of dealing with diversity in the classroom, and there has been a parallel growth in awareness for equity in the classroom. Ironically, there have been policies aimed at integration in place for about thirty years yet the potential for successful implementation of inclusion through good contextualisation is only just being realized. The thirty-year history of inclusion of disabled students into regular shows a slow beginning, with acceleration of this process occurring only very recently. Preparing regular teachers to deal with inclusion has been and continues to be a challenge. Preparing special education teachers for co-teaching and supportive roles in regular education settings is equally challenging. There have been numbers of more recent official recommendations, for example, the principles and policy objectives stated in the White Paper on Rehabilitation (HK Government, 1995), and the deliberations of the Board of Education Sub- committee on Special Education (1996), that encourage inclusion. However there are difficulties faced by regular teachers as they struggle with the concept of integration and they mostly arise through a lack of understanding of students with learning difficulties and a lack of strategies necessary to support these students in the regular class setting.

Patterning the integration of other countries

Historically, many countries tried to integrate disabled students into regular schools by placing them in special classes within the schools, and this has been attempted in Hong Kong. In most of these countries this move was followed relatively quickly by the integration of individual disabled students into regular classes. This was often followed by a backlash from regular teachers who felt unsupported and unable to help such students. In Hong Kong integration is taking place at a more measured and slower place. There are now more than sixty schools undertaking integration. Other policy reforms, for example school-based management, are introduced by way of a decree and a timeline. In fact up scaling of integration by small incremental steps is slowly taking place, and this tardiness may be a result of tensions between the Education Department and the powerful sponsoring bodies that manage and control schools in Hong Kong. Given the centralized and economically powerful position held by the Education Department of Hong Kong, it is a little surprising that reform is so slow.

The context of simultaneous reforms and unity of purpose

In attempting to implement inclusion, the context of reform in Hong Kong presents a difficulty in itself. Educational reform in Hong Kong has involved the rapid introduction of various policies, with an expectation that schools, through school-based management will have independence, in deciding for themselves, the priority of introduction. In that setting, inclusion may be given very low priority and its implementation could be delayed almost indefinitely. Additionally, the contextualisation of inclusion needs to have a unity of purpose. Recent experience in the observation of the integration process and as shown in the idealized model below, indicate that all elements, from the school's values through to parent involvement must be addressed. The authority charged with funding and policy implementation may not even demonstrate such unity. For example, within that authority, while one section is responsible for the implementation of the stated policy on integration in regular classrooms, and is making an effort to promote it, another section responsible for resources might be busily enforcing rulings on the requirement for a special education unit or class in each new school. Further need for contextualisation is evidenced by the unwillingness of schools to include students with learning difficulties. As previously discussed, early recommendations for integration in Hong Kong really came from overseas precedents, and it is suggested that far more attention to local needs and significant elements would have made the implementation of integration easier and quicker.

School cultural context and inclusion

Policy implementation or reform in Hong Kong often seems to experience cultural difficulties. Policies are drawn up, and publicized; yet schools are in general reluctant to take them up. Within the school, there are number of further potential cultural contextual barriers that should have been addressed. Classroom teachers without illumination may fear or resent students who are disabled or different. They may feel that they have enough "problems" to deal with, without the inclusion of "handicapped" students. Special schools may find it difficult to decentralize and integrate their services into regular schools. Inclusion has an uphill battle in the Hong Kong cultural setting, and in the school culture.

A pilot integration scheme involving more than sixty schools in which regular schools took in a small number of "integrators" with the incentive of receiving a bounty fund for each "integrator" is just beginning to show that schools may be becoming more accepting. In the scheme, This is evidence of a definite move to a more consistent approach, but it is still far from the concept of inclusion. While policy and context may be identified as key considerations in bringing about change, one very important factor is leadership. With effective leadership, aimed at the grass roots level of the school, school staff can become participatory teams. All staff can feel that they are part of the decision making in the school, and this is much more likely to lead to effective and appropriate change. Under such circumstances, there will be consensus on the schools aims and objectives, and this is most likely to come about when the school leaders and participants have a shared vision—This will give strength through unification of purpose. This is essential when it involves decisions that move a school towards a philosophy of inclusion.

In the Hong Kong context, there is increasing awareness by educators for the need for this type of educational reform. Current reform has targeted the need for dealing with diverse learners, parental involvement, student focus, a "no-loser" principle and a move away from syllabi that target examinations. (Education Commission, 1999). These types of reform initiatives can only benefit students who are included into regular schools.

However, the ability of leaders and educators to implement and support reforms remains as a further barrier to their full adoption and utilization. This resistance to change suggests that leaders and educators need two things. On the one hand they need to want to change, and on the other hand they need to feel that they have the skills to bring about change. Attitudinal change is one of the most difficult areas of human adaptation. Although schools do contain Resource Classes for "low achievers", these are carefully categorized and separated so as to preserve the status of the remainder of the school. Interestingly, since 2000, the Education Department has allowed resource teachers to work on a "non-withdrawal" in-class basis. Leaving the decision to schools though does not appear to have led to any marked change from the traditional Resource Class pattern. Enskilling people for change is a little easier. Skills to bring about change can be built up by for example, by training leaders in how to conduct workshops that target problem-solving, motivation techniques and support strategies. These can later be contextualised into programmes of staff development that will develop and harness those skills that enable staff to make decisions, to work collaboratively and to work in teams productively.

It seems that schools are never quite "ready" for inclusion, or have not thought about how to introduce inclusion into the current context. However, the reality is that if it were necessary to wait until everything was ready before beginning inclusion, it would never happen. In an idealized model there are needs in the environment that can be attended to that greatly facilitate the inclusion process, and the more of these that are addressed the more chance there will be of a smooth contextualisation of inclusion. The needs and facilitation elements in an idealized model are shown and discussed below. (Dowson, 2001).

Table 1
Facilitating Inclusion in Hong Kong Schools

Need	Specific Facilitation Element	General Facilitative Elements

• Inclusive school values and direction	• Leadership	Good policy implementation and continuing support
• Inclusion into society	• Publicity	Community consultation
• Acceptance by peers, teachers, and principals	• Whole school approach	School introspection
• Full classroom support	• Full funding	Good site-based strategic planning
• Competent and confident teachers	• Teacher education & support	Prioritization, task identification and implementation of ideas
• Planned initiation	• Careful preparation	Close liaison between policy-makers and implementers
• Acceptance by parents of regular students	• Contact and communication	Conversion of special education institutions to resource centers
• Early age inclusion	• School consultation	Shifts in centralized funding
• Parallel curricula	• Administration acceptance	Supportive legislation
• Included child preparation	• Parent guidance	Curriculum liberalization

In this facilitation model, inclusion reform entails consideration of a number of factors, all of which are important if optimal, and acceptable inclusion is to be attained. These factors include:

- Inclusion into society needs to be achieved through changing public attitudes toward the disabled. High quality and persistent publicity is one way this might be achieved.
- Acceptance by peers, teachers and principals requires the support of the "whole school approach." Acceptance by parents of regular students is particularly important and can be achieved by schools that, perhaps through School Management Initiatives, choose to address the issues and involve and inform parents in, for example, the move toward an inclusive school. The requirement in Hong Kong for the adoption by schools of School-Based Management makes this more likely.
- The inclusion class environment, in which identified regular students with negative attitudes about disabled students have been helped to adjust, is also important. A "Whole School approach" would include this activity, however, such an approach in Hong Kong appears to have involved much rhetoric, but little actuality.
- In the classroom, support for included students requires considerable funding, more than the bounty funding offered to schools in the Integrated School Pilot Scheme. This can be achieved by redirecting special education funding to disabled students who have been re-sited in regular schools. Another consideration is the regular classroom teacher. Teachers need to be confident about their ability to support the learning of disabled students. As previously discussed in relation to policy recommendations, this requires in-service and pre-service teacher education. Recent teacher education programmes address these issues, however programme participants are often disillusioned when they fail to find included students in regular classrooms. Yet to facilitate inclusion, regular teachers need to be able to carry out planned initiations in the classroom.
- Inclusion will generally not occur positively unless teachers make an effort to engage the class in activities that foster it. For example in a situation where a heterogeneous group is working on a separate project in which each group member needs to contribute, there would be an included student. Classroom factors include the flexibility of the teacher in helping the included child connect with the appropriate elements of the regular curriculum. Another factor that is helpful to the inception of inclusion is early commencement. The student who comes to school with peers from the pre-school is already an accepted member of the group.

Reform implementation is about contextualisation and facilitating change. As such, it should be characterized by resource-rich, incentive-driven, continuous support. Concomitantly at the individual school level, there is a need to develop good site-based strategic planning. Planning components include prioritization, task identification, ideas for implementation and community consultation.

Planning needs to be carried out in an atmosphere of school introspection. This is an occasion when a school takes a good hard look at its operation and comes up with a structured plan on how to implement reforms, including insights into problems that might occur. Implementation of policies is more likely to happen when there is closer liaison between those writing policy reform and those who have to implement it. This means there is a strong obligation for policy writers to lie out how reforms will be contextualised and implemented. In Hong Kong, there are examples of policies failing to "catch on" through a lack of contextualisation. The introduction of the less than radical Target Oriented Curriculum (TOC), eight years ago, has required research to find out how to conceptualize and facilitate its implementation, and this has led to it being implemented three times under a different name to gain acceptance, as teachers were so put off by the style of its implementation and apparent lack of local contextualisation.

Integration of disabled children into ordinary schools was a policy recommendation in 1986. Nothing significant occurred until research into how to conceptualize and implement it began eleven years later, in 1997. (Crawford et al, 1999). Policy reform in inclusion needs contextualisation through public education, and the conversion of categorical special schools, units and personnel to resource centers and people. Legislation and litigation about inclusion, such as that which occurred in America in the seventies, can facilitate reform. Such a "big stick" approach would probably go against the Hong Kong current reform toward school-based management.

However in Hong Kong, there is some "trickle-down" to the education system from recent law changes such as the

Disability Discrimination Ordinance (1998). Although this mainly targets employers, there have been recent reports of parents claiming that being refused admission to regular schools has discriminated against their disabled children. Equity has a long journey in Hong Kong. Another incentive to change is a shift in funding. There is rarely more money available, but there is no reason why the cutting of the educational funding pie cannot be altered. Special education groups, especially those with a self-serving interest, may be reluctant to share their resources through re-distribution.

There are other ways of sharing resources. For example, special schools by changing roles and devolving can place and support their students in regular education settings— the context of inclusion. Finally calls for curriculum reform may actually enhance reform in inclusion. The flexibility and tailoring in and of curriculum that is needed by included students may emerge, if the connection between curriculum reform and inclusion is made.

Contextualisation is an important factor for integration, and the growth toward eventual inclusion. Contextualising an issue within the local setting will make it more acceptable. One of the main objectives of education is to make quality of life better for all, and to achieve this the involved parties need to be able to see how it will happen in their context. In the case of inclusion, it has taken a long time for the importance of contextual factors to be taken into account when implementing change. In each of the schools which have begun to take in disabled students, there are, and will continue to be, individual stories of best practices and barriers that arise. (Crawford et al, 1999). Those who implement integration or inclusion must tread a careful path between the introduction of ideas that might be seen as too radical and non-local, and those ideas that are in context, but do little to enhance quality of life for all students, particularly those coming from an inequitable setting.

Conclusion

In Hong Kong, a number of recent steps toward education reform have been taken through efforts to decentralize education, raise standards, and increase accountability and professionalism. It has been argued that, for example, that developing the professionalism of principals must be contextualised prior to and at the implementation stage if a professional development programme is to truly meet their needs and increase the quality of education. Early signs suggest that, encouragingly, this is beginning to take place. Similarly in higher education, unrealistic demands and impractical expectations of the higher education curriculum, teacher educators and student teachers have failed to take into account the context and socialization environment in which education takes place—the schools. In addition there is a de-contextualisation effect created by reforms that contradict one another, e.g., standards- based assessment versus learner-centered teaching. Teachers entering schools for the first time need to be ready for the present, prepared for the future of reform and highly adaptable to change through contextually authentic teacher education. There is an indication that this is beginning to occur.

Another component of the movement toward quality education through reform has been the development of a Classroom Language Assessment test component for English language teachers. The test is contextualised in the sense that it an authentic assessment of English teachers in the classroom. It was partly contextualised in that stakeholder groups were consulted about it (the exception being the largest teacher union body in Hong Kong). It was also contextualised in an anticipatory manner by the government who put aside considerable funding, conjecturing that many teachers would want to enroll in language enhancement programmes. However, with regard to the timeframe and evolutionary nature of the policy, there has been less contextuality than might have been desirable, in that insufficient attention was given to test development, or to the amount of time that the total process of establishing the Classroom Language Assessment test would actually take in Hong Kong. Without good contextuality, the implementation of policy is liable to be delayed.

In the case of inclusion, implementation has been slow to materialize because of a lack of contextualisation. The current education reforms in Hong Kong contain a number of elements, such as curriculum adaptation and whole school approach that support the implementation of inclusion. However, contextualisation for successful inclusion needs to operate at all levels and involves complex interactions. There are however, indications of greater emphasis being placed on policy contextualisation at the implementation stage. Implementation of policies is generally more likely to be successful where closer liaison exists between those responsible for developing policy reform and those responsible for implementing it. At all stages there is, therefore a strong obligation for policy writers to consider contextualisation and to state clearly how reforms and policies have been contextualised. In Hong Kong, policies often fail to "catch on" through a lack of contextualisation.

In viewing contextualisation through the four areas of professional development of principals, teacher education, assessment of teacher classroom language assessment and inclusion, it is apparent that reform implementation has little chance of success unless significant contextualisation is undertaken by key players working closely together and make genuine attempts to communicate with each other during all stages of policy formation and implementation. There are indications that in some areas in Hong Kong that this is beginning taking place.

References

Au, A. (1998). Language standards and proficiency. In Asker, Barry. (Ed.) *Teaching language and culture: Building Hong Kong on education*, pp. 179-183. Longman: Hong Kong.

Apple, M.W. (2001). Markets, standards, teaching, and teacher education. *Journal of Teacher Education*, 52(3), 182-195.

Bickley, G. (1997). *The golden needle: the biography of Frederick Stewart* (1836-1889). David C. Lam Institute for East West Studies: Hong Kong.

Bodycott, P. (in press). Blurred visions on a landscape of reform: teacher education in Hong Kong. *Asia Pacific Journal of Education*. Singapore.

Board of Education. (1996). *Report of the Sub-committee on Special Education*. Hong Kong: Government Printer.

Chan, S. (1999). The Chinese learner—A question of style. *Education and Training*, 41(6/7), 294-304.

Cheng, Y.C. (2000). The characteristics of Hong Kong school principals' leadership: the influence of societal culture. *Asia Pacific Journal of Education*, 20(2), 68-86.

Cheung, M.B. (2000). The changing roles and needs of school principals in Hong Kong. In A. Walker, P. Begley & C. Dimmock (Eds.), *School Leadership in Hong Kong: A Profile for a New Century* (pp. 61-62). Hong Kong: Hong Kong Centre for the Development of Educational Leadership

Choi, Chee-cheong. (1998). Language standards: an HKEA perspective. In Asker, Barry. (Ed.), *Teaching language and culture: Building Hong Kong on education*, (pp. 184-192). Longman: Hong Kong.

Cochran-Smith, M. (2001). Editorial: Higher standards for prospective teachers—what's missing from the discourse? *Journal of Teacher Education*, 52(3), 179-181.

Coniam, D. & Falvey, P. (1999). The English language benchmark initiative: A validation study of the Classroom Language Assessment component. *Asian Pacific Journal of Language in Education*, 2(2), 1-35.

Coniam, D. and Falvey, P. (1996). *Setting language benchmarks for English language teachers in Hong Kong secondary schools*. Advisory Committee on Teacher Education and Qualifications: Hong Kong.

Coniam, D. (forthcoming 2002) Illustrating the demands of a performance test through an interactive multimedia syllabus. *International Association for Language Learning Technology Journal*.

Coniam, D. & Falvey, P. (forthcoming 2001). Awarding passes in the Language Proficiency Assessment of English Language Teachers: Different methods - varying outcomes. *Education Journal*. 29 (2).

Crawford, N., Heung, V., Yip, E. & Yuen, C. (1999). Integration in Hong Kong: Where are we now and what do we need to do? *Hong Kong Special Education Forum*, 2(2), 1-13

Disabilities Discrimination Ordinance. (1998). In *Law of Hong Kong*, Chapter 487. Hong Kong: Government Printer.

Dowson, C. (2001). Paper presentation to Making Integration Successful Programme, HKIEd, February 2001.

Education Commission (1992). *Education commission report no. 5*. Hong Kong: Government Printer.

Education Commission. (1996). *Education commission report No. 6*. (ECR 6). Hong Kong: Government Printer.

Education Commission. (1999). *Education Blueprint for the 21st Century: Review of Academic System: Aims of Education: Consultation Document*. Hong Kong: Government Printer.

Education Commission (2000). *Learning for life, learning through life: reform proposals for the education system in Hong Kong*. Hong Kong: Government Printer

Education Department (2002, February). *Continuing professional development for school excellence* (consultation paper on continuing professional development of principals). Hong Kong: Hong Kong Special Administrative Region of the People's Republic of China Education Department.

Falvey, P. & Coniam, D. (1997). Introducing English language benchmarks for Hong Kong teachers: a preliminary overview. *Curriculum Forum*, 6(2), 16-35.

Falvey, P. & Coniam, D. (2000). *Language benchmarks for lower secondary teachers of English: Findings and emerging issues*. In Issues in English Language Benchmarking. Proceedings of the Colloquium on English language benchmarking, (pp. 9-15). Hong Kong Institute of Education: Hong Kong Institute of Education.

Fullan, M. (1993). *Change forces: probing the depths of educational reform*. London: Falmer.

Gipps, C. (1995). What do we mean by equity in relation to assessment? *Assessment in Education*, 2(3), 271-281.

Greene, M. (1978). *Landscapes of learning*. New York: N.Y.: Teachers College Press.

Government of the Hong Kong Special Administrative Region (HKSAR). (2000). *Language benchmark assessment for teachers—English* *language: syllabus specifications, explanatory notes, specimen questions with suggested answers, scales and descriptors*. Hong Kong: Government Printer.

Helsby, G. (1999). *Changing teacher's work*. Buckingham, Pa: Open University Press.

Hong Kong Government (1995). White paper on rehabilitation. Equal opportunities and full participation: A better tomorrow for all. Hong Kong: Government Printer

Howey, K. R. (1998). *Introduction to the commissioned papers*. Information Series No. 376, (pp. 19-34). Columbus: ERIC Clearinghouse on Adult, Career, and Vocational Education, Center on Education and Training for Employment, the Ohio State University.

Lin, M.D. (2001, March). *Professional development for principals in Taiwan: the status quo and future needs trends*. Paper presented at the International Conference on School Leaders Preparation, Licensure, Certification, Selection, Evaluation and Professional Development, Taipei, Taiwan.

Lynch, R. (2000). *New directions for high school career and technical education in the 21st century*. Information series No. 384. Columbus, ERIC Clearinghouse on Adult, Career, and Vocational Education, Center on Education and Training for Employment, the Ohio State University.

Muta, H. (2000). Deregulation and decentralization of education in Japan. *Journal of Educational Administration*, 38(5), 455-467.

McTighe, J. (1997). What happens between assessments? *Educational Leadership*, 54(4), 6-12.

Sanaoui, R. (1999, February 2). *Defining certification standards for ESL instructors in Ontario*. Talk delivered to the Advisory Committee on Teacher Education and Qualifications. Hong Kong.

Walker, A. (in press). School leadership and management in the Asia Pacific region. In J. Keeves and R. Watanabe (Eds.), *The handbook on educational research in the Asia-Pacific region*. Kluwer Press: Netherlands.

Walker, A., & Dimmock, C. (in press). Moving school leadership beyond its narrow boundaries: developing a cross-cultural approach. In K. Leithwood & P. Hallinger (Eds.), *Second International Handbook of Educational Leadership and Administration*. Netherlands: Klewer.

Walker, A., Begley, P. & Dimmock, C. (2000). *School leadership in Hong Kong: A profile for a new century*. Hong Kong: Hong Kong Centre for the Development of Educational Leadership.

About the Authors

Chris Dowson
Senior Lecturer, Department of Educational Psychology, Counseling and Learning Needs
Hong Kong Institute of Education, Hong Kong.
Email: dowsonc@ied.edu.hk

Chris Dowson specializes in inclusion studies and communication. His research interests focus on hearing impairment, second language learning and quality indicators in education. He teaches courses on classroom practice and effective methods.

Peter Bodycott
Principal Lecturer, Department of English
Hong Kong Institute of Education, Hong Kong

Peter Bodycott has taught in schools and higher education in Australia, Singapore and Hong Kong. His main publications and research interests are in the areas of pre- service teacher thinking, second language learning and teaching, language immersion and the role of narrative in leadership and teacher education.

Allan Walker
Professor, Department of Educational Administration and Policy
The Chinese University of Hong Kong, Hong Kong

Allan Walker specializes in educational leadership and policy. His major research interests center on the influence of culture and context on educational administration and leadership, principal assessment and strategic planning in schools.

David Coniam
Professor, Department of Curriculum and Instruction
The Chinese University of Hong Kong, Hong Kong

David Coniam is a teacher educator, working with ESL teachers in Hong Kong secondary schools. His main publication and research interests are in language testing, computational linguistics and language teaching

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dom
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS

DIE/CINVESTAV
rkent@gemtel.com.mx   kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Education Policy Analysis Archives

Volume 11 Number 6 **febrero 10, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 2003, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Diferencias de resultados cognitivos y no-cognitivos entre estudiantes de escuelas públicas y privadas en la educación secundaria de Argentina: Un análisis multinivel

Rubén Cervini
Universidad Nacional de Quilmes, Argentina

Citation: Cervini, R. (2003, February 10.) Diferencias de resultados cognitivos y no-cognitivos entre estudiantes de escuelas públicas y privadas en la educación secundaria de Argentina: Un análisis multinivel, *Education Policy Analysis Archives*, *11*(5). Retrieved [date] from http://epaa.asu.edu/epaa/v11n6/.

Resumen

En este artículo se analizan las consecuencias de asistir a escuelas públicas o privadas sobre los logros cognitivos en Matemática y Lengua, y los resultados no-cognitivos, como las actitudes frente a la matemática y las "expectativas de éxito futuro' de los estudiantes del último año de la educación secundaria en Argentina. El estudio examina datos de más de 130.000 estudiantes en aproximadamente 3.300 escuelas, dependiendo del indicador de resultado que se trate, provenientes del Censo Nacional de Finalización del Nivel Secundario 1998. Los puntajes de matemática y lengua fueron basados en pruebas estandarizadas aplicadas a los alumnos al final del año lectivo. Para desarrollar el a1nálisis se utilizaron modelos lineales de niveles múltiples con tres niveles (alumno, escuela y Provincia).El autor encuentra que *(i)* la influencia relativa de las escuelas sobre los logros cognitivos es mucho más alta que la que ejerce sobre los resultados no-cognitivos; *(ii)* si se toma en cuenta la composición socioeconómica y cultural de la escuela, no hay diferencias entre escuelas públicas y privadas respecto al rendimiento en matemática, mientras que en Lengua existe una ventaja pequeña a favor del sector privado; *(iii)* el efecto más importante de la dicotomía público-privado se verifica respecto del nivel y distribución de la "expectativa de éxito futuro" del estudiante. Los resultados se discuten en términos de la teoría de la reproducción cultural y del proceso de segmentación del sistema educativo argentino. Se discuten algunas implicaciones políticas de los resultados obtenidos.

Differences in Achievement and Non-cognitive Outcomes Between Public-school and Private-school Students in Secondary Education in Argentina: A Multilevel Analysis

Abstract

In this article, effects of attending public or private schools on cognitive achievement (Mathematics and Language) and on non-cognitive outcomes (attitudes toward Mathematics and educational and success expectations) of students in the last grade of the secondary education in Argentina are explored by means of multilevel analysis. The study examined data of more than 130,000 students of approximately 3,300 high schools, depending on the outcome indicator considered, from the Censo Nacional de Finalización del Nivel Secundario 1998 (High School National Census of 1998). Student's math and language scores were based on standard tests applied to the students at the end of the academic year. Using multilevel linear modeling with three levels (student, school and state), the author found that *(i)* the relative influence of schools on cognitive achievements is much higher than on non-cognitive outcomes; *(ii)* there is no difference in Mathematics achievement between public and private schools once socioeconomic and cultural school composition ("peer group") are controlled for, while private schools have a small advantage over public school in Language achievement; *(iii)* the most important effect of the public-private dichotomous variable is detected in relation to the distribution of student "success expectations." The results are discussed in terms of both the "cultural reproduction theory" and the segmentation process of the educational system in Argentina. Some implications for policy are drawn from the analysis.

Introducción

Durante la década del noventa, bajo una clara hegemonía del discurso neoliberal, se intensificó la discusión en torno de la privatización del servicio educativo y de las posibles consecuencias según fuese el mecanismo para llevarla a cabo (Nota 1). A diferencia de épocas anteriores, los argumentos dominantes pro-privatización no fueron ético-ideológicos, sino más bien, pragmáticos, orientados por los criterios de "racionalidad administrativa y económica" (Zibas,1996). Se asumió que la escuela privada ofrecía, a más bajo costo, mayor calidad - más altos y mejores logros de aprendizaje -, en comparación con la pública. Ese supuesto implicaba aceptar que las diferencias efectivamente observadas entre los logros académicos de los estudiantes de las escuelas privada y publica no se debían exclusivamente a las probables desigualdades en la composición social del alumnado de cada sector, sino que, en gran medida, eran explicadas por ciertas características institucionales positivas de la escuela privada (Ej. estilo de gestión, autonomía administrativa y pedagógica, etc). Si esta hipótesis era cierta, entonces, se tornaba razonable proponer como forma de dinamizar y mejorar el sistema educativo, el traslado de tales características a la escuela pública (Tedesco,1992:12) o más radicalmente, ampliar la participación privada en la oferta educativa (Nota 2).

A pesar de la amplia difusión de la creencia en la superioridad académica del sector privado, no existían, hasta muy recientemente, investigaciones que, con datos estadísticamente representativos y tratamiento metodológico adecuado, la avalasen. La reciente disponibilidad de datos muestrales representativos, de gran cobertura, provenientes de los sistemas de evaluación de la calidad, han permitido el desarrollo de ejercicios de análisis metodológicamente más confiables. En Argentina, por ejemplo, se cuenta con algunos análisis comparativos sectoriales de los resultados en las pruebas estandarizadas de matemática y lengua de los alumnos del último año de la educación primaria (Cervini,1999; McEwan,2001). No existen sin embargo, investigaciones sobre el nivel secundario. Además, los estudios citados se concentraron en los resultados cognitivos, a pesar de que la mayoría de los análisis sobre las relaciones entre educación y sociedad realizados a los largo de los años noventa, han insistido en la importancia de los resultados escolares no-cognitivos en sí mismos, muchos de los cuales han sido incluidos explícitamente en los marcos curriculares nacionales como objetivos a lograr por el sistema educativo. Por otro lado, algunas investigaciones han demostrado que la eficacia institucional para conseguir logros cognitivos no está necesariamente correlacionada con la que se obtienen logros no-cognitivos (Knuver y Brandsma, 1993; Mortimore, 1988; Opdenakker y Damme, 2000), al igual que no lo están ambos resultados entre sí. Por eso, es importante que cualquier análisis comparativo entre sectores (Ej. público/privado) incluya ambos tipos de resultados escolares. Finalmente, los estudios existentes no han diferenciado las variaciones inter-institucionales ("entre-colegios") de las geográficas ("entre-Provincias"), llevando a un posible sobre-dimensionamiento de la primera.

Este trabajo se propone (i) investigar los efectos (a) de los colegios y Provincias y (b) del sector de dependencia (público/privado) sobre los logros cognitivos y no-cognitivos en el último año del secundario en la Argentina, y (ii) determinar si las posibles diferencias sectoriales (público/privado) se deben exclusivamente a la composición socioeconómica y a los recursos materiales institucionales. En otras palabras, se tratará de determinar si las diferencias sectoriales en los logros escolares dejan de ser significativas cuando se "controlan" los efectos de las variables socioeconómicas, demográficas y de recursos escolares. Los datos considerados provienen de las pruebas de Matemática y de Lengua, y de los cuestionarios del estudiante y del Director, aplicados a los estudiantes del último año del secundario (Bachillerato y Comercial) en el Censo Nacional de Finalización del Nivel Secundario de 1997 y 1998, realizado por el Ministerio de Cultura y Educación de la Nación de Argentina. Para el tratamiento de los datos se aplica el denominado "análisis estadístico de niveles múltiples" o modelos lineales jerárquicos (Aitkin y Longford, 1986; Bryk y Raudenbush,1992; Goldstein,1987).

En primer lugar, se discuten los aportes de la teoría de la "reproducción cultural" al entendimiento de los logros cognitivos y no-cognitivos y de la segmentación sectorial (público/privado) del sistema educativo (punto I). A continuación, se exponen antecedentes de investigación empírica en esta área (punto II), para posteriormente presentar algunas características de la educación privada en la Argentina (punto III). En el punto IV se precisan los objetivos específicos del trabajo, seguidos de los principales aspectos metodológicos (punto V). En el punto VI se analizan los resultados, para finalmente, extraer algunas conclusiones (punto VII).

I. Teoría de la reproducción y segmentación institucional

La "reproducción cultural" se realiza en todo espacio pedagógico "intra-institucional" y es afectada por la forma de diferenciación o segmentación "inter-institucional".

(i) Según la teoría de la "reproducción cultural", el niño hereda mediante la socialización familiar, cierto capital cultural acorde con su pertenencia de clase. El capital cultural socialmente más valorizado (dominante) tiene mayor probabilidad de aparecer entre las familias de las clases simbólicamente dominantes. El conjunto de códigos y comportamientos conexos que la escuela reconoce como legítimos y universaliza para el conjunto de la sociedad, es una variante, compatible y concordante, con ese capital cultural. La institución escolar realiza esa función legitimante

en la medida en que establece un "vinculo oculto entre aptitud escolar y herencia cultural" (Bourdieu,1977; Bourdieu y Passeron,1981). Dado que "las diferencias de aptitud son inseparables de diferencias sociales según el capital heredado" (Bourdieu, 1997:35), su uso y legitimación "tiende a mantener las diferencia sociales preexistentes". Al tiempo que el sistema escolar "separa a los poseedores de capital cultural heredado de los que carecen de él." (35), los fracasados son "condenados en nombre de un criterio colectivamente reconocido y aprobado, por lo tanto psicológicamente indiscutible e indiscutido, el de la inteligencia", y colocados en el extremo inferior de "una jerarquía única de formas de excelencia - dominada hoy en día por una disciplina, la matemática-." (44).

Además del capital cultural, "trabajo acumulado" capaz de conferir "fuerza, poder y beneficios a sus poseedores" (Bourdieu,2000b:131-35), Bourdieu introduce el concepto de *habitus*, esquemas de pensamiento, percepción, evaluación y acción, más o menos inconscientes (Bourdieu,1994:10), que al ser incorporado por cada individuo como miembro de una clase social, es decir, en condiciones determinadas de existencia, hace que los miembros de una misma clase sean lo más similar posible en el mayor número posible de respectos y lo más distinto posible de los agentes de otras clases (Bourdieu,1994:10).

El *habitus* implica predicciones -aspiraciones y expectativas -, "disposiciones respecto al porvenir constituidas a lo largo de una relación particular con un universo particular de cosas probables" (Bourdieu,1991:110), probabilidades definidas objetivamente por la sociedad", que determinan no sólo la práctica actual del agente sino también su relación práctica con el porvenir, ajustada a las condiciones sociales de su adquisición y realización. El *habitus* anticipa y contribuye a realizar un futuro probable y de esta manera, se hace "cómplice de los procesos que tienden a la realización de lo probable" (111). No es un ajuste conciente, "una evaluación exacta de sus probabilidades de éxito", sino más bien los resultados duraderos inculcados por las posibilidades e imposibilidades, "facilidades y prohibiciones que están inscritas en las condiciones objetivas", que a su vez, generan "disposiciones objetivamente compatibles" con ellas, que toman impensable las prácticas más improbables y llevan a "rehusar lo rehusado y a querer lo inevitable" (94), a excluir "todas las "locuras" ('esto no es para nosotros')

Bourdieu culmina igualando el concepto de habitus con el de capital cultural *incorporado*, "posesión que se ha convertido en parte integrante de la persona, en *habitus*" (B.2000b:140) y que puede adquirirse "de forma completamente inconsciente" (141), quedando "determinado para siempre por las circunstancias de su primera adquisición" (141). La adquisición familiar del capital cultural, diferente al escolar (Bourdieu, 1991, 1998a, 2000b), es "total, precoz, e insensible, efectuado desde la primera infancia", y genera un modo de relación "con la lengua y con la cultura que tiende a inculcar" que no puede ser igualada por el "aprendizaje tardío, metódico y acelerado". El *habitus* cuando es incorporado "naturalmente" se muestra como innato, fácilmente distinguible del "aprendido", y por lo tanto, artificial.

Discusión. Son pertinentes dos observaciones frente a estas tesis. Primero, la inclinación de Bourdieu a considerar al contacto con la cultura erudita y sofisticada, propia de la elite culturalmente dominante, reflejada en la valorización de las bellas artes (escuchar música clásica; asistir al teatro y a los museos), de un gusto refinado, de intereses y preferencias estéticas y de una forma de relación lingüística, como el mecanismo paradigmático a través del cual se establece la dominación cultural (Bourdieu,1988). Segundo, el concepto de *habitus* enfatiza el nivel inconsciente del agente social ("inconsciente manifestado en estado práctico"; esquemas que "funcionan en la práctica sin acceder a la representación explícita), próximo a la imagen de la corporización ("la forma en que una determinada cultura ha sido corporizada en el individuo"), atributo que debe ser también predicado respecto del capital cultural, dadala identificación entre ambos conceptos propuesta por el propio Bourdieu.

Estos supuestos constituyen obstáculos para entender la acción de la escuela o explicar las desigualdades reales en los resultados cognitivos y no-cognitivos. En primer lugar, la relación con una "cultura erudita" puede no ser tan preeminente o relevante en determinados contextos -por ejemplo, diferentes niveles del sistema educativo-, donde el peso de las aptitudes cognitivas parece innegable (Teachman,1987). En segundo lugar, al descalificar en su totalidad cualquier hipótesis que vincule el éxito en la escuela con la inteligencia, capacidad o aptitud del alumno que no sea consecuencia de una "construcción social", Bourdieu no atiende a las numerosas investigaciones que prueban la alta proporción de varianza en los resultados escolares explicada por aquel tipo de variables, aún después de "controlar" el efecto del origen social del alumno (Nash,2001). Finalmente, abundantes estudios empíricos han demostrado que las diferencias de aprendizaje entre escuelas también se sostienen después ejercer ese mismo "control" estadístico (Nota 3).

Para superar estos problemas, se ha propuesto ampliar el concepto de capital cultural (Nota 4). Además del *habitus* (nódulo central), el capital cultural comprende (i) "un conjunto de representaciones que sostiene el individuo en un momento dado de su vida" (Perrenoud, 1990: 52), las informaciones, los saberes, normas, valores, teorías, opiniones, creencias y proyectos de un futuro deseable, es decir, representaciones figurativas, conscientes, resultado de una construcción activa, y (ii) el grado de desarrollo intelectual o de la inteligencia operativa, la lógica natural, las habilidades procedimentales, la inteligencia, bases del "saber hacer", que además de haber afectado los procesos de aprendizajes anteriores, son puestos en práctica en los procesos actuales. Si en general, es correcta la tesis de que el *habitus* gobierna la práctica, en oposición al subjetivismo espontaneísta que sólo atiende a "las motivaciones surgidas instantáneamente de la libre decisión" (Bourdieu, 1991:95), debe reconocerse que al menos en parte, lo hace a través de las representaciones que sostiene el individuo, resultado de una construcción activa basada tanto en el *habitus* como en la acción reflexiva, a través de la cual el individuo "escoge, deforma, añade, en función de su personalidad, de sus posibilidades físicas y mentales." (Perrenoud,1990:48).

(ii) La segmentación institucional es *(a)* una respuesta de sectores sociales privilegiados para no perder sus posiciones relativas frente a la expansión de la escolarización (Bourdieu y Passeron,1981) o *(b)* el resultado de una estrategia de

alguna clase emergente para fundamentar su nuevo poder (Bourdieu,1989), ambos caracterizados por una diferenciación en las prácticas pedagógicas orientada a mantener o crear la distinción en los egresos del sistema.

Bourdieu (1989) ofrece un ejemplo paradigmático de construcción de un espacio institucional nuevo para nuevos conocimientos, ideología y práctica pedagógica, vinculado a la emergencia de un nuevo sector social. En su confrontación con la nobleza de espada y con la burguesía industrial y comercial, la nobleza de Estado asegurará su poder universalizando la valoración de la base de su poder: el capital cultural (Bourdieu, 1997:39) y la ideología del "mérito", llamada más adelante "competencia" (38), a través de los colegios (s.XVIII) y sus nuevas prácticas pedagógicas (Nota 5), semillas de los altamente selectivos colegios universitarios actuales.

La aparición o expansión, el fortalecimiento o la transformación de los colegios privados en países desarrollados, en cambio, ha sido una respuesta frente a procesos de democratización educativa. Caracterizados por exigir un alto nivel de "capital cultural" y por su capacidad de convertirlo en "capital escolar" altamente valorado, es decir, de alcanzar los más altos padrones de excelencia, los colegios privados legitiman la fragmentación, al tiempo que reproducen las diferencias anteriores a la democratización escolar (Bourdieu y Passeron, 1981). El proceso de diferenciación institucional de la educación secundaria en USA a inicios del siglo XX es un caso particular de este tipo de proceso. Tanto los colegios públicos como los Católicos experimentaron la presión de inmigrantes del área rural y de otras poblaciones externas, pero mientras los primeros adoptaron un modelo curricular flexible, atendiendo mejor a la gran variedad de aptitudes y preferencias de los estudiantes y reteniendo a sus alumnos hasta la graduación, los Católicos eligieron un curriculum académico estrecho, basado en la idea de la conveniencia de realizar fines académicos comunes (Nota 6). Esta diferenciación pedagógica y no la selectividad socioeconómica, explicaría por qué las escuelas Católicas gozaron de un mayor prestigio académico frente a las urbanas públicas. Entonces, la pregunta clave en este tipo de segmentación es si los colegios privados obtienen mejores resultados debido sólo a su composición socioeconómica homogénea y aventajada, o si, además de ello, adicionan el efecto de sus capacidades institucionales y pedagógicas.

II. Eficacia del sector privado: Antecedentes en la investigación empírica

Los logros cognitivos. A inicios de la década del ochenta, un estudio seminal sobre diferencias en logros cognitivos y en "aspiraciones para el futuro" (ambos considerados "resultados escolares") de alumnos que asistían a colegios secundarios de los sectores publico, privado católico y "otros privados" en USA, concluyó que las escuelas privadas producían mejores resultados cognitivos que las publicas y, en menor medida, más altas aspiraciones educacionales para el futuro (Coleman, Hoffer y Kilgore,1982). El hallazgo más fuerte de esta investigación y de otros análisis realizados posteriormente (Bryk, Lee y Holland,1993; Lee y Bryk, 1988) fue que "las diferencias sectoriales en la organización académica de la escuela es la principal razón de por qué los estudiantes de las escuelas Católicas logran niveles más altos y porqué el logro en esas escuelas es más equitativo entre los estudiantes de diferentes orígenes" (Lee et. al. 1998:317). La investigación y el debate sobre este tema han continuado en aquel país (Nota 7).

Los logros cognitivos en el Tercer Mundo. Durante la década de los años ochenta, el Banco Mundial realizó investigaciones para comparar logros académicos de escuelas secundarias públicas y privadas en 5 países del Tercer Mundo (Colombia, República Dominicana, Filipinas, Tanzania y Tailandia). Al sintetizar sus resultados, Jimenez y Lockheed (1995) concluyeron que, aún manteniendo constante los antecedentes del alumno y el sesgo de selección, "los estudiantes de las escuelas privadas se desempeñan mejor que los de las públicas en diversas pruebas de logro" (115). Sin embargo, otra revisión más crítica y focalizada sobre aspectos metodológicos claves de esas mismas y otras investigaciones, concluyó que las evidencias empíricas no indican una ventaja aplastante de las escuelas privadas sobre las públicas (Ridell, 1993: 384). En primer lugar, todos los estudios presentados por Jiménez y Lockheed (1995) usan el método de mínimo cuadrado ordinario (OLS), el cual modela "una proporción relativamente pequeña del total de la variancia en el rendimiento", no refleja la estructura de agrupamiento jerárquico de la realidad bajo investigación y es sumamente ineficiente para propósitos de ordenamiento de diferentes escuelas, defectos subsanables con el uso de la metodología de análisis de regresión multinivel (Nota 8) (Aitkin y Longford, 1986; Bryk y Raudenbush, 1992; Goldstein, 1987). En segundo lugar, y estrechamente relacionado al anterior, aún cuando se sabe que las características del "contexto socioeconómico" institucional poseen una fuerte capacidad explicativa de las diferencias 'inter-escuela' en los resultados escolares, esas variables no son explícitamente modeladas en todos los estudios (este es el caso de Filipinas, Colombia y Tanzania). El uso de un factor de corrección del "sesgo de selectividad" (Heckman,1979), propuesto para controlar el efecto de factores extraescolares no medidos (por ejemplo, aptitud y motivación), sin lo cual se sobreestimaría el efecto de la escuela como tal, no tiene la misma claridad y eficacia que las mediciones del contexto institucional ("peer group") (Nota 9). Cuando se usa el método de regresión multinivel y se modelan explícitamente variables del contexto socioeconómico, las diferencias público/privado desaparecen (Ridell,1993: Zimbabwe) o son "explicadas" casi totalmente por tales variables (Lockheed y Zhao,1991: Filipinas; Lockheed y Bruns,1990: Brasil). En tercer lugar, la variable criterio utilizada no es la misma en todos los estudios. En algunos casos se usa simplemente el rendimiento en algunas áreas (principalmente matemática y lengua), mientras que en otros se usa un promedio de logros o el valor agregado (Nota 10) en diferentes áreas. Otro problema que puede agregarse es el referido al tamaño de las muestras. Dado que el objetivo principal es comparar instituciones públicas y privadas, el número de ellas incluidas en el estudio se torna particularmente relevante. Sin embargo, la más extensa de las muestras incluyó 100 escuelas (Tailandia), mientras que otras son extremadamente pequeñas (Tanzania:13; Zimbabwe:29; Colombia: 35). Finalmente, observamos que en los estudios presentados no se incluye el análisis de cualquier indicador de resultado diferente al logro académico escolar.

Un estudio reciente sobre la diferencia público/privado en la educación primaria de Chile (McEwan,2001) merece una atención especial, principalmente por la magnitud de la muestra analizada. El autor compara los resultados en lengua y matemática de casi 150 mil alumnos del 8° año de la educación básica en 6 tipos de escuelas públicas y privadas.

Las estimaciones son ajustadas por indicadores socioeconómicos individuales y contextuales ("peer group") que muestran un gran efecto sobre las diferencias de rendimiento. Por ejemplo, la distancia inicial de las escuelas privadas (sin voucher) respecto de las publicas municipalizadas (aprox. 11 items) cae a menos de la mitad (4,84 items: Tabla 6). El autor detecta un importante efecto de las variables contextuales, aunque no es posible saber su importancia relativa en relación a las individuales (Nota 11). En los modelos ajustados, sólo las escuelas Católicas (con voucher) y las 'privadas sin voucher' continúan teniendo un efecto significativa. Sin embargo, cuando se introduce una versión modificada (policotómica) del factor de correción para el "sesgo de selección" de Heckman (1979), las diferencias residuales entre sectores desaparecen (Nota 12).

En Argentina, y respecto de la educación básica, los datos provenientes de las evaluaciones de la calidad educativa, realizadas por el Ministerio de Educación desde 1993, permitieron desarrollar los primeros análisis del tema en base a muestras extensas y representativas del país. Usando los resultados de las pruebas de Matemática e informaciones provenientes del cuestionario del alumno, del docente y del Director, Cervini (1999) encontró que las diferencias de rendimiento de los alumnos del 7° año en escuelas públicas y privadas, desaparecían cuando se tenían en cuenta el nivel económico y educativo y los bienes culturales familiares, la composición socioeconómica de la escuela ("contexto escolar") y la dotación de recursos materiales de la escuela (infraestructura y materiales didácticos). Las estimaciones se obtuvieron con el método de modelos lineales jerárquicos (multinivel) y fueron confirmadas con los datos de 1994, 1995 y 1997.

Un estudio más reciente y basado en los datos de 1997 (McEwan, 2001), evaluó la eficacia (rendimiento en matemática y lengua) de los diferentes tipos de escuela privada - católica subsidiada, subsidiada no-religiosa y privada no-subsidiada - en comparación con la escuela pública. Sus conclusiones fueron que, si se tienen en cuenta diversos indicadores socioeconómicos, los alumnos de las escuelas privadas no-subsidiadas no se diferencian de las escuelas públicas, tanto en sus logros de matemática como de lengua (Nota 13); las católicas subsidiadas tampoco son efectivas respecto de Matemática; las subsidiadas no-religiosas sí parecen ser diferentes, tanto en matemática como en lengua. Dado que estas últimas constituyen sólo el 1 % del total de alumnos (Carnoy, 2001:56), estos resultados son en general, coincidentes con las conclusiones del estudio anteriormente citado (Nota 14) (Cervini,1999).

Estas conclusiones sobre las diferencias público / privado, sin embargo, no pueden ser extrapoladas sin más a la educación secundaria y menos aún respecto de los estudiantes del último año de ese nivel, principalmente porque las composición socioeconómica de ambos sectores es notablemente diferente. Basado en los mismos datos y mediciones utilizadas en el presente trabajo, y aplicando también modelos jerárquicos lineales, Cervini (2002) constató que la probabilidad de aprendizaje en el nivel secundario está fuertemente vinculada a la institución educativa a la que se accede. Mientras que el 'capital económico familiar' determina las probabilidades de acceso y permanencia en la escuela secundaria, el 'capital cultural ' familiar y contextual (institucional) moldea el perfil de la distribución de los logros escolares en matemática y lengua. Y dentro de este acotamiento, más importante que el "volumen" de capital cultural 'incorporado' (Ej. años de estudios de los padres) es su activación (Lareau,1987), indicada indirectamente por la posesión de capital cultural "objetivado' relativo al sistema escolar. La incidencia de la 'herencia cultural familiar' del estudiante individual opera a través del agrupamiento en instituciones escolares. Los "culturalmente iguales' tienden a estar institucionalmente juntos y por tanto, sus diferencias de logro deben ser explicadas por otras características personales. Consecuentemente, los indicadores del 'contexto cultural institucional' muestran una capacidad predictiva del rendimiento notablemente mayor que sus homónimos a nivel del estudiante individual. En otras palabras, la injusticia educativa opera principalmente a través de la segmentación sociocultural del entramado institucional. El estudio mostró también el efecto notablemente alto de los antecedentes académicos (repitencia) del estudiante y del grupo, a tal punto que la 'proporción de repitientes' hace desvanecer el efecto del 'nivel educativo de los padres' como contexto, sugiriendo que el 'capital cultural incorporado' contextual ejerce su efecto a través de los logros escolares acumulados por el estudiante. Además, el estudio concluyó que los colegios difieren respecto del grado en que el origen sociocultural del estudiante incide en la distribución de los aprendizajes; algunos consiguen 'compensar' las desigualdades de origen de sus estudiantes más eficazmente que otros. Finalmente, se estableció que, luego de ajustar por todos los indicadores socioeconómicos, incluidos los culturales, la desigualdad 'inter-colegio' representa alrededor del 18% de la variación residual total y por lo tanto, su explicación debería buscarse en otros tipos de factores, algunos institucionales o pedagógicos, alterables y pasibles de ser distribuidos más equitativamente.

En ese estudio, sin embargo, no se incluyó la variable "sector de dependencia" (público / privado) del establecimiento, ni se consideró cualquier resultado 'no-cognitivo'. Tampoco fueron incluidas otras mediciones disponibles y relevantes para ajustar adecuadamente el efecto sectorial, tales como (la percepción de estudiante sobre) el involucramiento familiar educativo, la situación laboral del estudiante y la dotación institucional de recursos materiales, todo ellos insumos exógenos. El presente estudio pretende justamente llenar esos vacios.

Los resultados no-cognitivos. Casi todas las investigaciones discutidas anteriormente focalizaron su atención en los resultados cognitivos. Sin embargo, es comúnmente aceptado que la escuela también "produce" resultados no cognitivos. Aunque no muy abundantes, existen investigaciones que han evaluado el efecto diferencial de las unidades escolares sobre los resultados no-cognitivo, además de los cognitivos (Brookover et. al., 1979; Mortimore et.al.,1988; Knuver y Brandsma,1993). El primer aspecto a tener en cuenta sin embargo, es que los conceptos e indicadores utilizados varían notablemente. A veces se refieren a conductas escolares y se basan en registros administrativos, como la disciplina y la asistencia (Mortimore et. al., 1988) o en cuestionarios, como la integración social en la clase, las relaciones con los maestros (Opdenakker y Damme, 2000), el compromiso en el trabajo académico o los comportamientos escolares que pueden afectar el éxito en la escuela (Lee y Smith, 1993); en otras ocasiones, se trata de actitudes, intereses y auto-concepto respecto de diversas áreas curriculares y actividades de aprendizaje (Mortimore et. al., 1988; Opdenakker y Damme, 2000) y expectativas sociales más generales, como las "aspiraciones educativas para el futuro" (Coleman, Hoffer y Kilgore,1982). Un segundo aspecto se refiere a las marcadas diferencias entre las variaciones 'inter-escuela' de los resultados cognitivos y no-cognitivos. Mientras que dicha variación en los cognitivos se sitúa en torno de 20% en países desarrollados (Bosker y Witziers,1996) y no menos de 30% y hasta 60% en los del Tercer Mundo (Riddell,1997), en los no-cognitivos alcanza valores promedios en torno del 10%, aunque en algunos indicadores, la variación inter-escuela se torna prácticamente inexistente (Opdenakker y Damme, 2000; Mortimore et.al.,1988) (Nota 15). Estas diferencias de variación entre ambos tipos de resultados sugieren indirectamente, que no siempre existe concordancia en el grado de efectividad institucional para producirlos, ni alta correlación entre ellos. Finalmente, los resultados pueden variar de acuerdo al método de análisis utilizado y el nivel educativo de que se trate.

III. La educación privada en Argentina

Casi al inicio del presente siglo, Argentina mostraba un alto nivel de escolarización secundaria (Nota 16), comparado con la mayoría de los países de América Latina. Informaciones referidas a 1999 (SIEMPRO,2000) indican que la cobertura neta secundaria de la población de 13 a 17 años era del 82% (Nota 17). Sin embargo, la magnitud de ese logro se ve seriamente disminuida cuando se presta atención a indicadores que reflejan el grado de culminación efectiva de ese nivel educativo: sólo el 45,7% de la población de 18 a 24 año ha completado la educación secundaria. Por otra parte, ambos aspectos - escolarización y egreso - se distribuyen muy desigualmente por nivel socioeconómico familiar. Mientras que la escolarización de los jóvenes en familias situadas en el 1° quintil de ingreso ("el 20% más pobres") apenas sobrepasa el 70%, la del último quintil ("los más ricos") es superior a 97%. Además, de la población más pobre de 15 a 24 años que había comenzado el colegio secundario, casi el 30% lo abandonó sin terminarlo, situación que alcanza a sólo el 6,6% de la población más favorecida económicamente. Como consecuencia, el 24,4% de los jóvenes de 18 a 24 años en el 1° quintil ha completado la secundaria, porcentaje que llega al 76% entre esa población en el quintil de más alto ingresos. La reciente profundización de la crisis económica permite suponer que estos indicadores se pueden haber agravado. En resumen, en Argentina se constata un alto nivel de escolarización secundaria, acompañado con un igualmente alto grado de ineficiencia y segmentación socioeconómica.

La presencia del sector privado en la educación secundaria es manifiestamente alto. Según el Relevamiento Anual de 1999, realizado por el Ministerio de Educación de la Nación, el 27,3% de los estudiantes secundarios está matriculado en una unidad educativa del sector privado (Nota 18), el cual acumula casi el 25% del total de unidades educativas en el país. Este peso relativo del sector privado aumenta sostenidamente a medida que se avanza en los grados; en el último, casi el 35% de los estudiantes pertenece al sector privado (Nota 19). Como consecuencia, casi el 40% de los egresados del secundario proviene del sector privado.

La asistencia escolar secundaria muestra un evidente sesgo de selectividad socioeconómica. El 46% de los asistentes a establecimiento privados pertenece a familias situadas en los dos quintiles superiores de la distribución de ingreso *per capita* familiar, mientras que esa estimación apenas alcanza al 20% en el sector público (SIEMPRO,1999). Sin embargo, es importante advertir que, a pesar de ello, una parte considerable de los estudiantes en el sector privado - 30,7% - proviene de familias situadas en los dos primeros quintiles de la distribución de ingreso. Esta constatación es coincidente con el alto porcentaje de estudiantes en el sector privado (55%) que están matriculados en los establecimientos de la Iglesia Católica (Morduchowicz et.al.,1999), gran parte de las cuales atiende a poblaciones de bajos recursos.

La gran importancia del sector privado en Argentina es el resultado de un largo proceso histórico. Aparentemente, la expansión del sector aparece estrechamente relacionado con su progresiva desregulación y con el otorgamiento de subsidios con recursos públicos (Morduchowicz, 2001). El proceso de expansión de la matricula privada a partir de finales de los años cuarenta, cuando paralelamente se institucionaliza el subsidio y se acentúa la desregulación del sector privado (Nota 20), darían sustento a esta tesis. Caben dos observaciones. En primer lugar, el sector privado ya significaba poco más del 30% de la matrícula total. En segundo lugar, a partir de aquellos años es la tasa global de escolarización secundaria la que experimenta un crecimiento muy acentuado, reflejado en una notable aceleración de las tasas trienales de crecimiento relativo de la matrícula en ambos sectores (ver Gráfico 1). Inicialmente, es el sector público el que crece de forma más acelerada, pero desde finales de los cincuenta hasta 1970, el crecimiento relativo del sector privado pasa a ser siempre mayor que el público, hasta recuperar el techo de participación relativa en la matrícula total (cercana al 33%). El crecimiento 'rezagado' del sector privado daría soporte también a la hipótesis según la cual se expande, al menos en parte, como una reacción al proceso de escolarización universal. Persiste sin embargo, un interrogante clave respecto de la expansión de la matrícula privada: ¿cumplió la finalidad de mantener o inclusive, aumentar niveles de excelencia académica, o más bien, la de resguardar suficientes espacios institucionales socialmente homogéneos, puestos en jaque por la expansión de la escolarización? La pregunta es pertinente porque, hasta el inicio de su expansión, el sector privado no gozaba de un alto prestigio académico (Nota 21) ("formador de elite"). Durante las dos décadas posteriores, el crecimiento relativo del sector privado siempre fue inferior al del público, descendiendo como consecuencia, su participación en la matricula. Esta tendencia sólo parece revertirse durante los primeros años de la década del noventa, pero a partir de 1994 vuelve a descender notablemente.

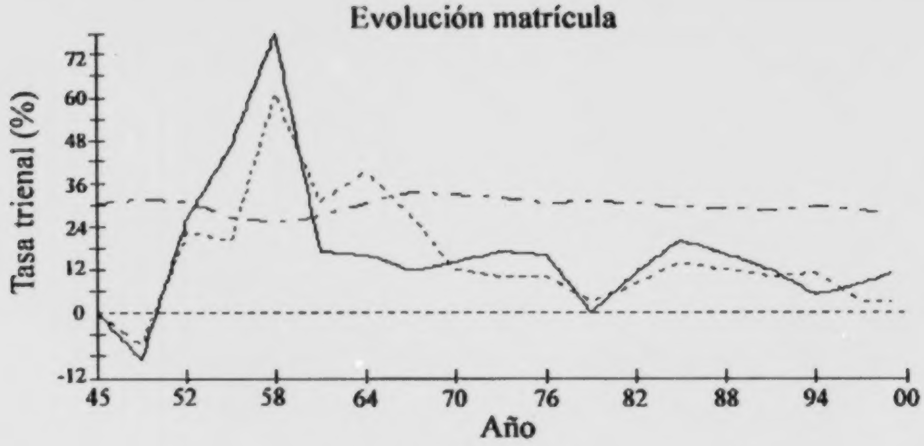


Gráfico 1

IV. Objetivos

En este estudio se aplica el análisis estadístico de multinivel (modelos lineales jerárquicos) para (i) investigar los efectos (a) de los colegios y Provincias y (b) del sector de dependencia (público/privado) sobre logros cognitivos y no-cognitivos de los estudiantes en el último año del secundario en la Argentina, y (ii) determinar si las posibles diferencias sectoriales (público/privado) se deben exclusivamente a factores de selección social (factores extraescolares o exógenos) y/o de recursos materiales institucionales. En otras palabras, se trata de determinar si las diferencias público-privado en los resultados institucionales promedios dejan de ser significativas cuando se introduce el "control" de factores socioeconómicos, demográficos y de recursos materiales.

La primera pregunta de investigación es relativa a la importancia del efecto de las diferencias entre los colegios en la determinación de los logros cognitivos y no cognitivos. Dada la relevancia de incluir y distinguir tanto los logros en conocimientos y habilidades cognitivas como los resultados no-cognitivos, el interrogante es si los colegios tienen igual influencia en ambos tipos de resultados educativos. La forma de responder esta pregunta es procesando, para cada resultado estudiado, un modelo "vacío" o incondicional (sin ningún predictor) que estima la magnitud relativa de cada uno de los componentes de la variación total, es decir, alumnos, colegios y Provincias.

La segunda pregunta de investigación es en torno al efecto de la diferenciación sectorial, es decir, si los estudiantes de los colegios privados consiguen, en promedio, resultados superiores a los de los públicos y si esa diferencia existe tanto en los resultados cognitivos como en los no-cognitivos. Para responder esta pregunta se adiciona a los modelos "vacíos" estimados en el paso inicial, la variable relativa al sector (Público/privado).

La tercer pregunta es si la probable diferencia sectorial se debe exclusivamente a la selectividad de entrada de los estudiantes y a la diferencia en la disponibilidad de recursos materiales entre los colegios de ambos sectores (público/privado), o si, por el contrario, otras características -educativas e institucionales- explicarían una proporción importante de esa hipotética diferencia. Entonces, se trata de determinar en qué grado la posible diferencia de logros entre los estudiantes 'público' y 'privados' se "explica" por la composición socioeconómica y la dotación de recursos institucionales.

V. Metodología: datos, variables y estrategia

Datos. Se analizan datos que provienen de (i) las pruebas de Matemática y de Lengua aplicadas a los estudiantes del último año del secundario (modalidad Bachillerato y Comercial), (ii) el Cuestionario del estudiante y (iii) el Cuestionario del Director, aplicados durante el Censo Nacional de Finalización del Nivel Secundario de 1998 (Nota 22), realizado por el Ministerio de Cultura y Educación de la Nación de Argentina. La evaluación se realizó a finales del año lectivo. Se incluyen solamente los estudiantes que dieron las pruebas de Matemática y/o Lengua, que hubiesen respondido el Cuestionario del estudiante y cuyo Director también hubiese respondido el cuestionario correspondiente. Los colegios en la modalidad técnica no son incluidos en el análisis, dado que las importantes diferencias curriculares respecto de las modalidades bachillerato y comercial impiden la comparación directa. Además, no se consideran a los estudiantes pertenecientes a colegios con información válida para menos de 10 estudiantes. Con estas condiciones, el archivo de Matemática queda constituido por 134.939 estudiantes en 2.708 colegios y el de Lengua por 131.066 estudiantes en 2.685 colegios. En ambos casos, alrededor del 42% de los estudiantes pertenece al sector privado, porcentaje muy próximo al de los egresados del secundario provenientes de ese sector.

Variables dependientes (individuales) (Nota 23). La principal característica del presente estudio es incluir variables criterio referida tanto a resultados cognitivos como no cognitivos. Los cognitivos se refieren a sendas pruebas estandarizadas de Matemática y de Lengua, ambas con 40 ítem de opción múltiple. Para los no-cognitivos se utilizan cuatro indicadores; dos de ellos son mediciones exploratorias de actitudes hacia la matemática (conocimiento escolar): motivación para la Matemática (MOTIVACION) y valoración de la Matemática (VALORACION); los otros dos miden actitudes socioeducativas más generales: aspiración educativa (ASPIRACION), definida como planes del estudiante para el futuro inmediato, y expectativa de éxito (ÉXITO), definido como auto-predicción de éxito futuro en diferentes áreas.

Variables independientes: Todas estas variables están definidas a nivel del aula o del colegio (Nota 24)("contexto institucional"). A excepción de las variables provenientes del Cuestionario del Director, originalmente referidas al colegio, y de la dependencia del colegio (público/privado), todas las restantes expresan promedios o proporciones de características individuales de los estudiantes en el colegio. Las variables independientes pueden agruparse en subconjuntos:

- sector de dependencia (PRIVADA), variable "dummy", siendo Público la variable base;.
- nivel socioeconómico familiar: compuesto por capital económico, medido por la disponibilidad de bienes de uso durable y servicios en el hogar (BIENES:) y capital cultural: nivel educativo del padre y de la madre (EDUCACIÓN), disponibilidad de libros (LIBROS) y disponibilidad de libros, fichas y apuntes escolares (DIDÁCTICOS) en el hogar;
- implicación ("involucramiento") familiar en educación, referido a (la percepción del estudiante sobre el) interés de la familia en las actividades escolares (INTERES FLIA) y el plan para el futuro inmediato del estudiante (ASPIRACION FLIA);
- características personales: aquí se incluyen variables de significados diferentes:
- alumno repitiente (REPITIENTE), variable "dummy"; variable base: alumnos no-repitiente. En este tipo de estudios es habitual "controlar" por los logros anteriores del estudiante ('insumo de entrada') o algún indicador *proxy* de este concepto. En nuestro caso, la única medición disponible es el antecedente de repitencia del estudiante, indicador que expresa en parte, su capacidad o aptitud académica (Nota 25);
- la cantidad de horas que el estudiante dedica a trabajar (HS_TRABAJO), no sólo aporta mayor precisión a la medición del origen social del estudiante, sino que permite además, detectar variaciones en la "oportunidad de aprender" (Carroll,1967), uno de los principales condicionantes del nivel de rendimiento; para nivel primario en Argentina, ver Cervini, 2001;
- género (FEMENINO), variable "dummy"; variable base: masculino; se incluye porque la mayoría de las investigaciones indican que los varones alcanzan mejores logros en Matemática y las mujeres en Lengua;
- los recursos institucionales: incluye el estado de la infraestructura del establecimiento (INFRAESTRUCT) y los recursos didácticos disponibles en el establecimiento (DIDÁCTICO_COL), ambos provenientes del Cuestionario del Director.

El comportamiento, la eficacia y el significado conceptual de las mediciones del nivel socioeconómico familiar ya han sido probados en un estudio anterior (Cervini, 2002). Las mediciones referidas a actitudes o percepciones del estudiante deben considerarse "exploratorias". Todas las variables han sido estandarizadas con la finalidad de tornar directamente visible (en proporciones) los cambios que se producen en la varianza no-explicada del rendimiento y el peso relativo de cada una de las variables en la explicación de la variación del rendimiento.

Técnica de análisis. Para el análisis de las relaciones entre cada uno de los diferentes indicadores de resultado (cognitivos y no-cognitivos), por un lado, y el conjunto de variables de insumo consideradas, por el otro, se utilizó la técnica de "análisis estadístico por niveles múltiples" (Aitkin y Longford, 1986; Bryk y Raudenbush,1992; Goldstein,1987). Esta es una técnica correlacional adecuada para analizar variaciones en las características de los individuos (Ej. rendimiento en matemática) que son miembros de un grupo (Ej. colegio) que a su vez, hace parte de otra agregación (Ej. Provincia), o sea, mediciones que forman parte de una estructura agrupada y jerárquica. La técnica permite la descomposición de la variación de una variable (Ej. rendimiento) en sus componentes "dentro del grupo" (intra-colegio; intra-Provincia) y "entre grupo" ('inter-colegio'; 'inter-Provincia') y el análisis de la asociación entre variables en esos diferentes niveles de agregación. El modelo se compone de una Parte Fija, con los parámetros que definen una línea promedio para todos los estudiantes de todos los colegios de todas las Provincias, y de una Parte Aleatoria, que muestra en cada nivel de agregación, la estimación de la variación de los parámetros determinados en la parte fija. La principal ventaja de esta técnica es que modela simultáneamente los diferentes niveles de variación (por ejemplo, estudiante, colegio y Provincia), permitiendo, por tanto, saber qué proporción de la variación del rendimiento escolar se debe principalmente a características del estudiante, del colegio y de la Provincia. Para estimar la probabilidad del efecto de las variables se usa el *test* de la razón de máxima verosimilitud (Nota 26).

Estrategia metodológica. Los datos permiten definir modelos con tres niveles de agrupamiento: el estudiante (*nivel 1*), el colegio (*nivel 2*) y la Provincia (Nota 27) (nivel 3). El ordenamiento secuencial del análisis adoptado responde a los objetivos propuestos y a la técnica de análisis: se comienza con la partición de la varianza de cada indicador de resultado, por nivel de agregación (1° pregunta de investigación) y a continuación, se modelan las distancias de los resultados entre los sectores público y privado (2° pregunta); finalmente, se adicionan los indicadores referidos a distintos aspectos del contexto institucional (3° pregunta). Dado el objetivo principal del análisis, el criterio para determinar el modelo final correspondiente a cada indicador de resultado (cognitivos y no-cognitivos) no es sólo el nivel de significación de su efecto, sino que además, debe producir una disminución significativa del efecto de (PRIVADA), es decir, de la distancia entre los logros promedios esperados de los estudiantes de los sectores público y privado. Finalmente (4° paso), se determina el grado de variación de los indicadores de resultado por sector de gestión (público/privado) y se intenta develar la existencia de posibles interacciones entre (in)equidad educativa y tales sectores.

VI. Resultados y análisis

Estadísticas descriptivas. Los datos se comportan de acuerdo a lo previsto. Inicialmente, sin ninguna clase de "control", los estudiantes del sector privado consiguen más altos logros en Matemática y en Lengua (71,4 y 72,5%, respectivamente), que los del sector público (59,2 y 61,1%, respectivamente). Los rendimientos promedio de los colegios se comportan de forma similar. La expectativa de éxito y la aspiración educativa también tienden a ser más alta (en más del 10%) en los estudiantes de colegios privados; en cambio, no hay diferencia sectorial en las dos actitudes específicas frente a la Matemática. Los coeficientes de variación indican nítidamente que el sector privado es más homogéneo que el público en cuanto a los resultados cognitivos y las actitudes generales, mientras que en las actitudes referidas específicamente a la Matemática, la diferencia es más suave (Cuadro 1)

1 - Media y coeficiente de variación de los indicadores, por sector de dependencia

Indicadores	Media		C.V.	
	Pub.	Priv.	Pub.	Priv.
Matemática	59,2	71,4	.336	.246
Lengua	61,1	72,5	.301	.226
Motivación	140	141	.298	.294
Valoración	92,5	93,3	.225	.194
Éxito	115	129	.309	.243
Aspiración	4,40	4,95	.355	.263
Matemá_c	56,1	70,0	.242	.174
Lengua_c	58,2	71,3	.198	.151
Motiva_c	142	142	.089	.089
Valora_c	93,3	93,7	.068	.061
Éxito_c	115	128	.109	.098
Aspira_c	4,28	4,92	.124	.096
bienes	7,61	9,34	.317	.276
educ_flia	7,07	8,89	.404	.328
libros	3,02	3,77	.454	.326
didactic	2,27	2,37	.353	.244
%repite	32,9	13,1	.014	.026
%femenino	61,2	62,1		
hs_trab	1,21	0,66	.710	.753
Flia_int	3,32	3,43	.237	.208
Flia_des	3,36	3,50	.259	.220
Infraest	143	174	.217	.072
Rec_Dida	11,2	14,1	.318	.192

Todas las mediciones relativas al nivel socioeconómico y cultural de la familia indican que los estudiantes de las escuelas privadas provienen, en promedio, de hogares más favorecidos. El indicador de tenencia de libros y fichas escolares, sin embargo, no presenta una diferencia tan nítida. Ello debe ser adjudicado principalmente, al tipo de medición utilizado, con reducida amplitud y cantidad de valores. A pesar de ello, el sector público también es más heterogéneo. Los indicadores relativos a la percepción del estudiante sobre el interés y las aspiraciones educativas familiares se comportan de la misma forma que los anteriores. La proporción de repitentes es notablemente menor en los colegios privados, reflejando el comportamiento "expulsivo" del sector. Pero, por otra parte, ese sesgo de selectividad es más heterogéneo que el observado en el sector público; la explicación podría residir en la existencia de unidades escolares muy diferentes - unas "contenedoras" y otras "expulsivas" - dentro del mismo sector privado. Es también notable la capacidad de diferenciación sectorial conseguida por el indicador relativo a la situación laboral del estudiante: la intensidad laboral en el sector público duplica a la del privado. Además, los recursos institucionales son más abundantes y se distribuyen más homogéneamente en el sector privado (Cuadro 1).

En general, los resultados escolares cognitivos y no-cognitivos están débilmente asociados. Las correlaciones entre ambos resultados cognitivos y entre las dos actitudes

2 - Ceficientes de correlación simple (r-Pearson) entre indicadores de resultados					
Resultados	Len gua	Moti vación	Valori zación	Aspi- ración	Éxito
Matemática	.528	.165	.098	.178	.212
Lengua		.069	.085	.224	.281
Motivación			.375	.167	.193
Valorización				.174	.275
Aspiración					230

específicas frente a la Matemática son las más fuerte. La "expectativa de éxito" se asocia, aunque moderadamente, con todos los otros (Cuadro 2). Estas estimaciones confirman la relevancia de investigar ambos tipos de resultados escolares.

Análisis de regresión. Los resultados y la discusión se presentan de acuerdo a los pasos definidos en el punto sobre estrategia de análisis.

1° paso: El efecto "colegio" (modelo incondicional). La variación "inter-colegios" de todos los indicadores es estadísticamente significativa (Nota 28). En consecuencia, los colegios difieren entre sí respecto de los resultados que obtienen sus estudiantes, sean cognitivos como no cognitivos. Sin embargo, es sobre los primeros que la institución educativa tiene su mayor efecto. A ella es atribuible alrededor del tercio de la variación total de ambos rendimientos (Nota 29). En cambio, esa misma variación es de 11% cuando se trata de la expectativa de éxito futuro o de la aspiración educativa, y de 5% en el caso de la motivación y la valorización de la matemática. Por otra parte, sólo los logros cognitivos difieren significativamente entre Provincias. Como consecuencia, las variaciones "entre-estudiante" de los 'no-cognitivos' alcanzan valores muy superiores a los observados en los cognitivos. En términos generales, esta distribución es razonable ya que los resultados cognitivos le deben más a la escuela que los no-cognitivos, vinculados más estrechamente con el origen y contexto familiar (Nota 30) (Cuadro 3, Modelos "vacío").

En resumen, si bien el efecto del agrupamiento institucional de los estudiantes es evidentemente más fuerte en los resultados cognitivos, los no-cognitivos referidos a las actitudes socioeducativas generales también muestran una relación importante con ese criterio de agrupamiento. Entonces, los estudiantes obtendrán más altos o bajos rendimientos, aspiraciones educativas y expectativas de éxito, dependiendo del colegio al que asisten, predicción muy atenuada cuando se refiere a las actitudes específicas sobre la matemática, un conocimiento esencialmente escolar. De esta forma, se ha respondido a la primer pregunta de investigación.

2° paso: El efecto sector (PRIVADA). Todas las estimaciones de la variable "dummy" (PRIVADA) al ser introducida en cada uno de los modelos incondicionales, muestran un efecto estadísticamente significativo, es decir, los estudiantes de los colegios privados obtienen resultados, tanto cognitivos como no-cognitivos, significativamente más altos que los de colegios públicos. En los dos logros cognitivos y en los no-cognitivos generales (éxito y aspiración), la introducción de (PRIVADA) produce una disminución relativa importante en la variación 'inter-colegio' (alrededor del 20% y 32%); el rendimiento en Matemática, por ejemplo, cae 6,8 puntos porcentuales (de 0,341 a 0,273). Este comportamiento no se observa en las actitudes académicas de los estudiantes (motivación y valorización de Matemática). Puede concluirse que las diferencias existentes entre colegios respecto de estas actitudes, le deben muy poco a la distinción entre público y privado, y por lo tanto, las excluiremos del análisis subsiguiente (Cuadro 3, Modelos 'Privada').

La dicotomía público/privado también produce una disminución importante de la variación 'inter-provincia', es decir, un indicador definido a nivel de colegio produce efectos sobre las desigualdades de resultados en un nivel de agregación superior, la Provincia (Nota 31). Esto sugiere que las Provincias no son exactamente idénticas respecto de su composición público / privada y que una parte substancial de la variación entre ellas se relaciona con tal diferencia de composición. Es importante notar que si en el modelo no se hubiese especificado el nivel Provincia, esta variación habría sido imputada incorrectamente al nivel 'colegio' (Cuadro 3).

3- Estimaciones de modelos incondicionales ("vacío") y con efecto de (PRIVADA) para los resultados cognitivos y no-cognitivos.												
Nivel y Privad	Resultados cognitivos				Resultados no-cognitivos							
	Matemática		Lengua		Expectativa de éxito		Aspiración educativa		Motivación matemática		Valorización matemática	

	Vacío	Priv.	Vacío	Priv.	Vací	Priv.	Vací	Priv.	Vací	Priv.	Vací	Priv.
Prov	0,131	0,105	0,106	0,077	0,027	0,025	0,015	0,010	0,004	0,005	0,008	0,009
Col	0,341	0,273	0,330	0,259	0,111	0,075	0,111	0,071	0,052	0,051	0,045	0,043
Estad	0,565	0,565	0,618	0,618	0,875	0,876	0,895	0,895	0,950	0,950	0,954	0,954
PRIV	--	0,565	--	0,578	--	0,394	--	0,427	--	0,056	--	0,088
Nota: Prov: provincia; Col: colegio; Estud: estudiante; PRIV: (PRIVADA)												

3º paso: La "explicación" de las diferencias público-privado. Las constataciones anteriores justifican la prosecución del análisis, orientado ahora a identificar los factores que explican las diferencias entre colegios de los sectores público y privado. Para ello, las variables disponibles se introducen en los modelos que contienen el término (PRIVADA).

Los resultados obtenidos indican que la diferencia de logros cognitivos entre las escuelas públicas y privadas ha dejado de ser estadísticamente significativa, con mayor evidencia en Matemática que en Lengua. La estimación de (PRIVADA) para Matemática en el modelo incondicional (=0,565) cae a 0,028 cuando se incluyen las variables de ajuste (Cuadro 4). De aquí se infiere que la superioridad inicial, no ajustada, de los rendimientos promedios de colegios privados se explica por las diferencias en la composición socioeconómica de entrada y en la disponibilidad de recursos materiales institucionales. A igualdad en esas variables de insumo (o "control"), los colegios públicos obtienen los mismos resultados promedio que los privados.

No sucede lo mismo con uno de los indicadores clasificados como no-cognitivos generales. Los estudiantes a punto de egresar de un colegio secundario privado poseen una 'expectativa de éxito' muy superior a la de sus colegas del sector público, aún cuando se haya ajustado por las mediciones referidas a la composición socioeconómica y a la dotación de recursos materiales del colegio. Esta conclusión, sin embargo, no es aplicable a la 'aspiración educativa', ya que en este caso la diferencia público/privado desaparece cuando las estimaciones son ajustadas (Cuadro 4).

Más allá de que la expectativa de éxito es una medición más confiable que la de aspiración educativa (5 ítem en la primera frente a un único ítem en esta última), es muy probable que ambas capten aspectos psicoeducativos diferentes. La estructura de ambos modelos es diferente: mientras todos los indicadores de las características familiares poseen efectos significativos sobre la aspiración educativa, no sucede así con la expectativa de éxito. Ambas son principalmente herencia o transferencia familiar, pero también el sistema educativo parece tener un efecto propio que aunque leve, altera la distribución original de la expectativa de éxito, al tiempo que se limita a reproducir la aspiración educacional. Adviértase también que la variación 'residual' de (ÉXITO) continúa siendo significativo - poco más del 4,3% de la varianza residual total. Esto sugiere que existen otros factores, probablemente institucionales, que afectan la distribución de este valor social.

4- Estimaciones del efecto de (PRIVADA) y de las variables de "control", y variación residual por niveles, en cada indicador de resultado.

Indicadores institucionales	Resultados cognitivos		Resultados no-cognitivos	
	Matemática	Lengua	Expectativa de éxito	Aspiración educativa
PRIVADA	0,028*	0,071**	0,117***	0,025****
(error std)	(0,025)	(0,022)	(0,014)	(0,010)
bienes	--	--	0,073	0,026
educación	--	--	--	0,072
libros	0,211	0,198	--	0,084
didácticos	0,030	0,071	0,072	0,016
%repitientes	-0,134	-0,139	-0,077	-0,030
%femenino	--	--	--	--
hs_trabajo	-0,069	-0,261	--	0,021
interés flia.	--	--	0,071	0,021
aspiración flia	0,030	--	--	0,112
infraestruct.	0,024	--	--	0,016

didactico_c	0,022	–	0,024	–
Provincia	0,050	0,022	0,006	0,002
Colegio	0,178	0,163	0,043	0,011
Estudiante	0,565	0,618	0,876	0,895
Test	31300,1	315350,1	368130,5	369279,9

(*) no significativo.
(**) Si se extrae PRIVADA, el *test* es 315360,2; diferencia con 315350,5 = 10,1 que, con un parámetro adicional, tiene probabilidad de ocurrencia = 1.48 por mil, o sea, superior al criterio de selección de 1 por mil.
(***) Prob. ≤0.001.
(****) Si se extrae PRIVADA, el *test* es 369285,7; diferencia con 369279,9 =5.8 que, con un parámetro adicional, tiene probabilidad de ocurrencia = 1.6%, o sea, no significativo si el criterio de decisión es de 1 por mil.

4º paso: Variaciones e interacciones sectoriales. Nuestro interés es determinar si existe algún efecto de interacción sólo entre el sector y las variables definidas a nivel de colegio (*nivel 2*). Queda por tanto, excluido el análisis de posibles efectos de interacción con las características individuales. Antes que nada, parece conveniente comparar sectorialmente la importancia relativa de la variación inter-colegio de cada uno de los indicadores de resultado (Nota 32). Para ello, recalculamos los modelos 'vacíos' de cada variable criterio para ambos sectores. Los resultados indican que, en general, la importancia relativa de las desigualdades en los resultados promedios institucionales tiende a ser mayor en el sector privado; sin embargo, es en Matemática (cognitivo) y en expectativa de éxito (no-cognitivo) donde las diferencias son más notorias (Cuadro 5). Restringiremos el análisis de interacción, entonces, a estos dos indicadores de resultados.

5 - Partición de la varianza en cada variables criterio, por sector de gestión, según niveles de agregación.

Nivel	Matemática		Lengua		Expectativa de éxito		Aspiración educativa		Motivación Matemática		Valorización Matemática	
	Públic	Priv.	Públic	Priv.	Públi	Priv.	Públi	Priv	Públi	Priv.	Públi	Priv.
Prov	0,117	0,147	0,083	0,070	0,030	0,021	0,014	0,013	0,006	0,005	0,090	0,011
Col	0,271	0,345	0,250	0,266	0,058	0,104	0,065	0,086	0,048	0,054	0,041	0,049
Estud	0,631	0,605	0,695	0,512	0,925	0,888	0,930	0,916	0,953	0,947	0,957	0,949

Los resultados del análisis de interacción (Nota 33) respecto del logro en Matemática indican que la repitencia es el factor con un efecto diferencial más nítido, es decir, en la escuela privada la distancia entre los rendimientos promedios esperados de los repitentes y los no-repitentes es significativamente mayor que en el sector público. Respecto de la repitencia, entonces, la escuela pública distribuye más igualitariamente el saber escolar de matemática. Pero, por otro lado, los indicadores de "capital cultural" tienen mayor fuerza en el sector público, particularmente los recursos didácticos (significación: 6 por mil). Respecto de los recursos institucionales no se pueden extraer conclusiones definitivas: mientras existen indicios de que en el sector público la infraestructura pesa más (1,3%), en el privado los recursos didácticos parecen ocupar esa posición (7,7 por mil) (Cuadro 6).

En cuanto a la "expectativa de éxito", el sector privado aparece, en general, más inequitativo que el público. Los dos indicadores de "capital cultural" tienen un mayor efecto en el sector privado, al igual que la situación laboral del estudiante. La repitencia aparece nuevamente como un factor de diferenciación sectorial - más fuerte en el privado. El efecto de la percepción del estudiante sobre el interés de su familia y los recursos institucionales tienen, estadísticamente, el mismo peso en ambos sectores.

6 - Estimaciones del efecto de las variables de "control", por resultado y sector.

Indicadores	Matemática		Éxito	
	Públic.	Priv.	Públic.	Priv.
libros	0,230	0,154	ns	0,076**
didacticos	0,037*	ns	0,047	0,080*
%repitentes	-0,094	-0,184*	0,051	-0,071**
hs_trabajo	-0,068	-0,081	ns	-0,032**
Interés flia			0,078	0,059
infraestruct.	0,032*	ns	ns	ns

didáctico_c	ns	0,037*	0,020	0,031

(**) Prob. ≤0.001; (*) Prob. ≤0.01. *Test* de máxima verosimilitud.

VII. Discusión y conclusiones

Una de las principales características de la investigación educativa en las dos últimas décadas ha sido la notable proliferación de los estudios con el enfoque de "eficacia escolar", preocupados por develar los factores que explicarían las diferencias observadas en los logros promedios de las instituciones educativa. Es posible identificar cuatro clases de factores que explican la distancia entre el resultado promedio de un colegio cualquiera y el promedio general de todos los colegios: (i) la composición del colegio (contexto institucional), (ii) los factores socioeconómicos exógenos (individuales del alumno), (iii) las políticas y las prácticas institucionales y (iv) los efectos no medidos de esa escuela. Se ha propuesto denominar a este efecto global, que suma sin distinguir esos 4 tipos de factores, como "tipo A" (Raudenbush y Willms,1995). El efecto "tipo B", en cambio, se refiere solamente al efecto específico de las políticas y las prácticas institucionales y a los efectos no medidos del colegio y se define como la distancia entre el promedio obtenido por el colegio y el 'promedio esperado' por el conjunto de colegios que son similares respecto de su composición (contexto institucional) y de las características socioeconómicas exógenas de sus estudiantes. Entonces, mientras que en el efecto tipo A se incluyen factores que están más allá del control de maestros y Directores, en el tipo B se trata de estimar el efecto que expresa el aporte institucional específico. El 'efecto A', efecto total, es de interés del estudiante y de su familia: asistir al colegio que obtenga los mejores resultados, sin importar sus causas y de acuerdo a sus posibilidades económicas familiares. El 'efecto B', en cambio, es del interés del gobierno, de los Directores, de los docentes, y se vincula a la demanda por una mayor eficacia educativa, ajustada a principios de justicia distributiva en educación.

El principal objetivo de este trabajo ha sido estudiar la posible incidencia de la segmentación sectorial (público/privado) de la educación Media desde la perspectiva del 'efecto B', incluyendo resultados tanto cognitivos como no-cognitivos. El análisis de los datos ha permitido establecer las siguientes constataciones:

- existe una importante variación en los promedios institucionales de los resultados cognitivos y no-cognitivos en torno del promedio global. Asistir a un colegio determinado no es irrelevante, sino que implica una cierta probabilidad de obtener un resultado por encima o por debajo del promedio general. Por lo tanto, el agrupamiento de estudiantes en colegios tiene un efecto ('tipo A') sobre los resultados, reflejado precisamente en esa variación. Consistente con la literatura internacional, el efecto del sistema educativo sobre los resultados cognitivos (Matemática y Lengua), alrededor de un tercio de la variación total, es mucho mayor que el ejercido sobre los no-cognitivos. Entre estos últimos, el efecto sobre las actitudes educativa generales (expectativa de éxito futuro y aspiración educativa) es superior al ejercido sobre las actitudes relacionadas a la Matemática - 11% y 5%, respectivamente.

- una proporción significativa de esa variación está asociada a la dicotomía público/privado: los resultados promedios obtenidos por los colegios privados son superiores a los de los colegios públicos. Esta afirmación es válida para ambos resultados cognitivos y para las actitudes educativas generales (aspiración educativa y expectativa de éxito futuro), pero no se sostiene para los no-cognitivos vinculados específicamente al conocimiento de la Matemática. Desde el punto de vista de las familias ("demanda"), entonces, la creencia de que el colegio privado produce mejores resultados no puede rechazarse. Desde la perspectiva de la política educativa, por el contrario, el problema es determinar qué proporción de esa ventaja se debe a peculiaridades institucionales, y cuánto a factores socioeconómicos contextuales.

- la superioridad de los colegios privados en ambos resultados cognitivos desciende abruptamente cuando se introduce el "control" de los indicadores referidos a la composición del establecimiento y sus recursos materiales (edilicios y didácticos). En Matemática, la diferencia se desvanece totalmente, mientras que en Lengua, persiste un leve ventaja, no significativa al 1 por mil. Entonces, respecto del aprendizaje curricular, la institución privada ejerce poco o ningún 'efecto tipo B', es decir, la explicación de la ventaja inicial del sector privado ('efecto A') no debe buscarse en las políticas y las prácticas institucionales ni en los 'efectos no medidos', sino principalmente en las determinaciones exógenas del propio sistema educativo, más allá del control de docentes y directivos.

- el colegio privado potencia la expectativa de éxito, es decir, los síndromes actitudinales relacionados a la auto-estima, al auto-concepto o al sentido de autoeficacia general. Sobre esta dimensión el colegio privado consigue diferencias nítidas: de dos colegios socialmente similares, es más probable que el privado consiga mejores resultados que el público. Si bien es cierto que el 'efecto colegio' sobre este resultado no-cognitivo (11%) es muy inferior al de los cognitivo (33%), no debe olvidarse que la medición utilizada para la 'expectativa de éxito' es exploratoria y no tan confiable como las pruebas cognitivas. Es razonable hipotetizar que el uso de mediciones más precisas y confiables para las dimensiones no-cognitivas acentuaría la tendencia registrada con los datos analizados.

- la distribución de la expectativa de éxito tiende a ser más inequitativa en el sector privado que en el público, es decir, más dependiente del capital cultural familiar y de la situación laboral del estudiante. Por otro lado, el comportamiento de los datos relativos a Matemática, más oscilantes, no permiten extraer alguna conclusión terminante a respecto de la (in)equidad sectorial.

- Finalmente, los indicadores utilizados han explicado la mitad de la variación 'inter-colegio', aún resta por explicar la otra mitad. Otras características, seguramente vinculadas a 'políticas y prácticas institucionales', pero que poco o nada tienen que ver con la dicotomía publico/ privado, están en la base de esas desigualdades inter-institucionales residuales y deben ser objeto de futuras investigaciones.

En una revalorización de *La Reproducción* de 1989, Bourdieu (1998) afirma que "todas las pruebas de la metodología empirista" (128), producidas por investigaciones realizadas en Estados Unidos y Reino Unido, son confirmatorias de la teoría de la reproducción. Las altas correlaciones entre capital cultural y rendimiento académico constatadas en este estudio también la apoyan. Pero, con esa misma "metodología empirista", se verifica también que una porción importante de las desigualdades en los resultados cognitivos en la educación secundaria debe se explicada por otros factores. Si bien una porción significativa de las desigualdades entre colegios confirma la transformación del capital cultural en capital escolar, otra de magnitud similar, podría deberse a características personales del alumno o al 'efecto escuela'. Los componentes no-cognitivos, como el 'sentido de éxito', también responderían a este razonamiento, aunque en forma menos pronunciada. Es decir, la predicción de "una correlación muy estrecha entre las *probabilidades objetivas* científicamente construidas y las *esperanzas subjetivas*" (Bourdieu, 1991:94) parece mejor ajustada al comportamiento de los datos.

El sistema de educación secundaria argentina está profundamente segmentado. La población socialmente más aventajada concurre al colegio privado. Desde el punto de vista de esas familias, la elección es racional: los colegios privados consiguen mejores resultados cognitivos y no cognitivos que los públicos. Desde el punto de vista político, la respuesta no es la misma. Si por un lado, existen indicios de que a igualdad socioeconómica, las escuelas privadas consiguen mejores resultados no cognitivos generales, por el otro, los datos cuestionan la supuesta 'eficacia institucional' del colegio privado en cuanto a logros cognitivos. La segmentación público-privada en Argentina ha significado el resguardo de espacios institucionales más orientados a reforzar y potenciar las diferencias de *habitus* que la distancia entre los componentes cognitivos del capital escolar. Los alumnos de dos colegios del mismo nivel socioeconómico, uno público y el otro privado, muy probablemente conseguirán similares resultados cognitivos. La principal o exclusiva causa de diferencias cognitivas se debe a factores socio-culturales (composición institucional) y a la disponibilidad de recursos materiales (edilicios y didácticos) en la institución educativa. El carácter regresivo del subsidio a la educación privada con recursos públicos, detalladamente demostrado por Morduchowicz (2001), tiende ciertamente a reforzar esa tendencia, en detrimento de las poblaciones más carenciadas. Adicionalmente, no existen indicios de una mayor equidad en el sector privado en comparación con el público. En resumen, la segmentación institucional cumple su papel reproductivo, pero no logra legitimarse agregando aprendizaje al esperado de acuerdo a la distribución del capital cultural en la sociedad.

Los colegios del sector privado gozan de "mayor flexibilidad y autonomía institucional" y consecuentemente, de un mayor "dinamismo y capacidad de adaptación a las necesidades de su matricula", desarrollado al amparo de su históricamente creciente desregulación por parte del Estado (Morduchowicz,2001; Narodowski, 2001). Pero no es evidente que estas características tengan que ver con las diferencias en los resultados educativos, es decir, no existe base empírica para creer que la extensión de tales características institucionales al sector público incidiría significativamente en una mejora de los logros educativos.

Notas

1. Para el caso Argentino, ver por ejemplo, Ciccioni (1999) y Cosse (1999).

2. Para una crítica de este argumento, ver Carnoy (2001).

3. Para la discusión actual sobre "eficacia escolar" ver: Teddlie y Reynolds (2001) y Thrupp (2001a; 2001b).

4. En Bourdieu, la relación entre capital cultural y *habitus* es ambigua y oscilante, conduciendo interpretaciones contradictorias. Mientras que para algunos, en el habitus "no todo es espontaneidad o automatismo", aunque sólo parece contener una forma de conciencia parcial, lagunosa, discontinua (García Inda,2000:27), para otros la sustancia del *habitus*, "... el *habitus* en si mismo, aún en sociedades tribales no diferenciadas, está hecho de capital cultural o, en el sentido más amplio de la palabra, de conocimiento (incluidas las habilidades)." (Lash, 1993:197).

5. El corte institucional educativo "instituye unas fronteras sociales análogas" a las existentes en la sociedad, y se hace patente no sólo en "el contenido y (...) en la organización del trabajo" escolar, sino también en la misma vida del estudiante; los centros de excelencia se distinguirán por sus marcos de comportamiento muy estrictos, formas de aprendizaje muy escolares y "sobre todo un ambiente de urgencia y competición" (Bourdieu, 1997:35).

6. Ver en Lee, V. et.al. (1998) bibliografía sobre la historia de ese proceso.

7. Raudenbush y Bryk (1986) reanalizaron los mismo datos con la técnica de modelo jerárquico, llegando a la conclusión de que las diferencias entre los rendimientos promedios de ambos sectores se explicaba totalmente por las diferencias en la composición socioeconómica de ambos, mientras confirmaba a las escuelas Católicas como más equitativas. Para algunas referencias actualizadas sobre la investigación empírica y el debate en esta área en USA, ver McEwan (2001).

8. Ridell (1993) informa los resultados obtenidos al aplicar ambas metodologías a los datos de Zimbabwe. Mientras

que el método OLS indica que las escuelas más eficaces eran dos tipos de escuelas privadas (confesional y elitista), con el análisis multinivel, que modela adecuadamente la covariancia dentro de los agrupamientos, las publicas y "otras privadas" resultaron más eficaces. Ver también el ejemplo de la nota anterior.

9. Más allá de que la situación dominante en los países del Tercer Mundo, no es la "elección" de escuela por los padres, sino la profunda "segmentación" socioeconómica institucional del sistema educativo, la alta complejidad del método de Heckman no compensa sus resultados, especialmente cuando se dispone de mediciones adecuadas del nivel socioeconómico del grupo de alumnos. Jimenez et.al. (1991), principales promotores del método Heckman, observan que si se incluyen variables socioeconómicas "contextuales" en el análisis, los valores de los términos de selectividad "se hacen insignificantes, indicando que las variables del grupo de alumnos ("peer group") podrían capturar sus efectos", sobre cuya base concluyen que "adicionar variables del grupo podría sustituir la engorrosa técnica de corrección de los sesgos de selectividad" (405).

10. Si bien el "valor agregado" tiene una mayor fuerza conceptual, no está libre de problemas metodológicos. El análisis de su comportamiento a lo largo de un período considerable de tiempo ha permitido constatar su alto grado de inestabilidad. Esto tiene una especial relevancia cuando se trata de evaluar el grado de eficacia institucional. En consecuencia, el uso del "valor agregado", calculado en base a una prueba al inicio y otra al final del año lectivo, no ofrece necesariamente mayor confiabilidad en las conclusiones y debería ser considerado como un criterio útil y exploratorio para evaluar los niveles relativos de rendimiento de los alumnos.

11. El autor evalúa el efecto de las variables grupales (nivel escuela) estimando la diferencia del efecto del tipo de escuela cuando se procesa el modelo completo (ajuste por variables individuales y grupales) y el modelo que contiene sólo las variables individuales. Sin embargo, no informa sobre el resultado inverso, es decir, del modelo que contiene sólo las variables grupales.

12. Dado que en este estudio se utiliza sólo el método de regresión ordinario, queda la duda de si el tratatamiento de los datos con el método de regresión multinivel, con modelamiento explícito del nivel escuela, no habría tornado prescindible el factor de correción.

13. El autor también incluye la repitencia como indicador de eficacia. Sin embargo, consideramos que este indicador, tal cual es medido en el cuestionario del alumno, es inadecuado para tal fin. En el cuestionario, al alumno de 7° grado se le pregunta si alguna vez repitió de grado. Dado que las mayores tasas de repitencia se producen en el 1° y 2° grado de la primaria y que la tasa de cambio de escuela es importante, la probabilidad de que el alumno repitente provenga de otra escuela donde efectivamente repitió, es bastante alta. La invalidez del indicador se acentúa cuando se tiene en cuenta además, el hecho de que existen escuelas "expulsoras" de repitentes (la mayor parte en el sector privado, aunque también existen en el sector público), que pasan así su "pasivo" a las escuelas "receptoras" de esos alumnos.

14. Es importante apuntar algunas diferencias metodológicas que pueden explicar las no coincidencias menores. En primer lugar, Cervini (1999) aplicó la técnica de modelos jerárquicos. En segundo lugar, existen algunas diferencias importantes en la construcción de las variables independientes de "control" que podrían influir sobre la "eficacia" para captar sus correlaciones con el rendimiento. Así por ejemplo, para la medición del nivel económico familiar, Cervini (1999) construyó un índice de nivel económico que aprovechaba la información de los 17 bienes de uso durables en el hogar incluidos en el cuestionario, mientras que McEwan sólo utiliza 3 de esos ítem, en forma de variables "dummy".

15. Por ejemplo, Mortimore et.al. (1988) encuentran las siguientes variaciones: asistencia escolar, 5,6%; autoconcepto, 8,4% y actitudes frente a la matemática, 12,2%. En Opdenakker y Damme (2000), la variación 'inter-institucional total (escuela+aula) es de alrededor de 5% en motivación y auto-concepto, y de 10% en integración social y relaciones con el maestro (pp.175). Ambos estudios aplican análisis de multinivel.

16. La Ley Federal de Educación (N° 24.195) de 1993 estableció la Educación General Básica obligatoria de 9 años, conformada por la exPrimaria (7 años) más los dos primeros años de la exSecundaria. El ciclo posterior (tres años), denominado Polimodal, posee las siguientes orientaciones: humanística, social, científica y técnica. Sin embargo, la implementación de este ciclo ha sido muy desigual entre las provincias. En el presente trabajo, el marco de referencia es la exSecundaria (5/6 años), incluyendo todos los años equivalentes de la antigua y nueva estructura del sistema educativo.

17. Con este porcentaje, la Argentina en un punto intermedio entre la escolarización secundaria en América Latina y el Caribe, próxima al 60%, y en los países desarrollados (100%). Ver Cuadro 4-1 en Ocampo (2000).

18. Esta estimación es consistente con lo informado por el SIEMPRO para 1999: el 25% de la población escolarizada en el secundario asiste a establecimientos del sector privado.

19. 1°:23,6%; 2°:25,6%; 3°:28,2; 4°:31,0; 5°:34,3. Elaboración propia sobre la base de informaciones de la Red Federal de Información Educativa del MCyE.

20. Durante ese periodo 1947 einicio de los sesenta, se promulga la ley 13047 (1947), que "sistematizó e institucionalizó el subsidio estatal a la educación de gestión privada", se eliminan los exámenes finales obligatorios para el sector privado y se legitima su autonomía (Morduchowicz,2001).

21. Los colegios secundarios de mayor prestigio eran en general, los dependientes de las Universidades Nacionales (públicas).

22. El Cuestionario del Director se aplicó en 1997 y en 1998. Sin embargo, en este último año una parte importante de los cuestionarios fue respondida posteriormente al día del operativo de evaluación y posteriormente, enviada por correo. Esta circunstancias arroja dudas sobre la calidad de los datos recolectados, además de una caída importante de la cobertura efectiva. Por eso, se decidió utilizar los cuestionarios de 1997, aplicados de acuerdo al procedimiento standard establecido. Cuando el Director no hubiese respondido el cuestionario en 1997 y sí en 1998, se utilizó este último (323 colegios).

23. Mayor detalle sobre la construcción de las variables está disponible con el autor.

24. Dado que el objetivo es evaluar el efecto del sector (la distancia público/privado) - variable definida a nivel colegio -, se espera que las variables individuales en sí mismas no tengan efectos significativos. Esta hipótesis se contrastó y confirmó con los datos. Los resultados no se exponen en honor a la simplicidad de lectura.

25. Está bien probado en la literatura no sólo la estrecha correlación entre repitencia y nivel socioeconómico, sino también su efecto "estigmatizante" negativo. Pero, dado que en el presente estudio, su efecto es "controlado" por todos los indicadores socioeconómicos, se asume que el efecto residual propio expresa en gran parte, condiciones aptitudinales del alumno.

26. Bajo hipótesis de nulidad de diferencia igual a 0 (cero), la diferencia entre valores de máxima verosimilitud de dos modelos sigue la distribución de chi-cuadrado con grados de libertad (gl.) igual al número de nuevos parámetros

27. El nivel Provincia no es el foco de atención de este estudio. Su inclusión se justifica porque de esta forma, se obtiene un estimación más precisa de la variación "entre-colegio". De no hacerlo, ésta última variación sería sobreestimada. Para el análisis, la provincia de Buenos Aires se divide en Gran Buenos Aires (Conurbano) y resto de la Provincia

28. Los errores estándar correspondientes y los valores del *test* de máxima verosimilitud para cada modelo pueden ser solicitados al autor.

29. Estos resultados son convergentes con los presentados en un trabajo anterior (Cervini,2002)

30. También es posible que una parte de las diferencias se deba atribuir a las diferencias en las propiedades de los indicadores de resultados cognitivos (pruebas estandarizadas) y no-cognitivos utilizados. Sin dudas, estos últimos están más expuesto a "contaminación por el contexto inmediato". De cualquier forma, las magnitudes obtenidas parecen estar en intervalos razonables, si se comparan con estimaciones obtenidas con la técnica de multinivel y en base a escalas de medición con alto grado de confiabilidad (ver nota 15)

31. Teóricamente, se espera que el factor afecte principalmente la variación del rendimiento en su mismo nivel. Así, las diferencias entre los rendimientos promedios de los colegios se deberían explicar principalmente por características que son comunes a todos los estudiantes de un mismo colegio (variables grupales). De la misma forma, la suma de las distancias entre los rendimientos obtenidos por los estudiantes y el rendimiento promedio del colegio al que pertenecen, deberían explicarse principalmente por las variables personales (individuales) del estudiante

32. Adviértase que se trata del peso relativo con respecto a la variación total, concepto diferente al de la variación expresada por el coeficiente de variación, y que fuera analizada al inicio de este apartado.

33. El procedimiento seguido es el siguiente: (a) el término interactivo se crea multiplicando el indicador por la variable "dummy" (PRIVADA); (b) cada término interactivo se incluye, uno por vez, en los modelos del Cuadro 4; (c) se determina la significación estadística con el *test* de máxima verosimilitud.

Bibliografía

Aitking, M. y Longford, N. (1986). Statistical modelling issues in school effectiveness. *Journal of the Royal Statistical Society* A, 149, pp. 1-42.

Bosker, R.J. y Witziers, B. (1996). *The magnitude of school effects or: Does it really matter which school a student attends?* Paper presented at the Annual Meeting of the American Educational Research Association

Bourdieu, P. (1977). Cultural reproduction and social reproduction. En: *Power and Ideology in Education*. Ed. J. Karabel and A.H.Halsey. New York: Oxford University Press.

Bourdieu, P. (1988) *La Distinción. Criterios y Bases Sociales del Gusto*. Madrid: Taurus Ediciones.

Bourdieu, P. (1989). *La Noblesse d'État. Grandes école et esprit de corps.* Paris: Minuit.

Bourdieu, P. (1991) *El sentido práctico*. Madrid: Taurus Humanidades.

Bourdieu, P. (1994) ¿Qué es lo que hace una clase social? Acerca de la existencia teórica y práctica de los grupos. *Revista Paraguaya de Sociología*. Año 31, N°89, pp.7-21.

Bourdieu, P. (1997). *Razones prácticas.* Barcelona: Anagrama, Colección Argumentos.

Bourdieu, P. (1998) *Capital cultural, escuela y espacio social*. Madrid: Siglo XXI Editores. Cap 8, pp. 125-128.

Bourdieu, P. (2000a). Sobre el poder simbólico. En: Andrés García Inda (Org.): *Poder, Derecho y Clases Sociales*, Cap. II. Bilbao: Desclée de Brouwer.

Bourdieu, P. (2000b). Las formas del capital. En: Andrés García Inda (Org.), *Poder, Derecho y Clases Sociales*, Cap. IV. Bilbao: Desclée de Brouwer.

Bourdieu, P. y Passeron, C. (1981). *La reproducción. Elementos para una teoría del sistema de enseñanza.* Barcelona:Laia.

Bryk, A., Lee, V. y Holland, P. (1993). *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.

Bryk, A. y Raudenbush, S. (1992). *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods.* Newbury Park, CA:SAGE.

Brookover, W., Beady, C., Flood, P., Schweitzer, J. y Wisenbaker, J. (1979). *Schools, social systems and student achievement - schools can make a difference*. Nueva York: Praeger.

Carnoy, M. (2001). La privatización ¿mejora las escuelas? *Propuesta Educativa*, 24, pp. 53-57.

Carroll, J.B. (1963). A model of school learning. *Teachers College Record*, N°. 43, pp. 723-733.

Cervini, R. (1999). *Calidad y Equidad en la Educación Básica de Argentina*. Factores Asociados al Logro Escolar, N° 5. Buenos Aires: Ministerio de Cultura y Educación de la Nación.

Cervini, R. (2002). Desigualdades Socioculturales en el Aprendizaje de Matemática y Lengua de la Educación Secundaria en Argentina -Un modelo de tres niveles. *Revista Electrónica de Investigación y Evaluación Educativa* (http://www.uv.es/RELIEVE). (Dictamen favorable para publicar en el "Número en construcción" actual.)

Ciccioni, A. (1999). El movimiento de las charter schools: una amenaza y una oportunidad para la educación pública argentina. *Propuesta Educativa*, 20, pp. 66-74.

Coleman, J. Hoffer, T. y Kilgore, S.(1982). *High School Achievement: Public, Catholic and Private Schools Compared*. New York: Basic Book.

Cosse, G. (1999) El sistema de voucher educativo: una nueva y discutible panacea para América Latina. *Propuesta Educativa*, 20, pp.51-65.

García Inda, A. (2000). Introducción. La razón del derecho: entre habitus y campo. En Andrés García Inda (Org.) *Poder, Derecho y Clases Sociales*, (pp.9-60). Bilbao:Desclée de Brouwer.

Goldstein, H. (1987) *Multilevel Models in Educational and Social Research*. London: Griffin.

Heckman, J. J. (1979). Sample selection bias as a specification error, *Econometrica*, 47, pp. 153-161.

Jimenez, E., Lockheed, M.E., Luna, E. y Paqueo, V. (1991). School effects and costs for private and public schools in the Dominican Republic. *International Journal of educational Research*, 15, pp. 393-410.

Jimenez, E. y Lockheed, M. E. (1995). Public and private secondary education in developing countries: A comparative study, *World Bank Discussion Paper* 309. Washington: The World Bank

Knuver, A. y Brandsma, H. (1993). Cognitive and affective outcomes in school effectiveness research. *School Effectiveness and School Improvement*, Vol.4 (2), pp. 189-204.

Lash, S. (1993). Pierre Bourdieu: Cultural Economy and Social Change. En: Craig Calhoun, Edward LiPuma y Moishe Postone (Ed.) *Bourdieu- Critical Perspectives*. Chicago: The University of Chicago Press.

Lareau, A. (1987). Social class differences in family-school relationships: The importance of cultural capital. *Sociology of Education*, vol. 60 (2), pp. 73-85

Lee, V. y Bryk (1988). Curriculum tracking as mediating the social distribution of high school achievement. *Sociology of Education*, 61, pp. 78-94.

Lee, V., Chow-Hoy, T., Burkam, D., Gerverdt, D. y Smerdon, B. (1998). Sector difference in high school course taking: a private school or catholic school effect?. *Sociology of Education*, Vol. 71 (October), pp. 314-335.

Lee V. and Smith, J. (1993). Effects of school restructuring on the achievement and engagement of middle-grade students. *Sociology of Education*, Vol. 66 (July), pp. 164-187.

Lockheed, M.E y Bruns, B. (1990). School effects on achievement in secondary mathematics and Portuguese in Brazil. *World Bank Staff Working Paper* N° 525. Population and Human Resources Department. Washington: The World Bank.

Lockheed, M.E. y Zhao, Q. (1991). The effects of local control on student achievement and attitudes in secondary education in the Philippines. *World Bank Staff Working Paper*, Population and Human Resources Department, Washington: The World Bank.

Morduchowicz, A.(2001). *(Des)Regulación y financiamiento de la educación privada en la Argentina*. Buenos Aires (trabajo no publicado).

Morduchowicz, A., Marcón, A., Iglesias, G. Andrada, M. Perez, J. Campán, V. y Duro, L (1999) *La educación privada en la Argentina: Historia, Regulaciones y Asignación de Recursos Públicos*. Buenos Aires (no publicado)

Mortimore, P., Sammons, P., Stoll, L., Ecob, R. y Lewis, D. (1988) The effects of school membership on pupils' educational outcomes. *Research Papers in Education*, Vol. 3 (1), pp. 3-26.

McEwan, P. (2000).The effectiveness of public, catholic, and non-religious private schools in Chile's voucher system, Stanford University (mimeo)

McEwan, P. (2001) Educación pública y privada en el Cono Sur: Un análisis comparativo entre Argentina y Chile. Trabajo presentado al Seminario Reforma Educativa en el Cono Sur: Logros y tareas pendientes. BID. Washington.

Narodowski, M. (2001). Segregación socioeconómica, regulaciones y privatización en el sistema educativo argentino: el caso de la provincia de Buenos Aires. *Propuesta Educativa*, 24, pp.41-52.

Nash, R. (2001) Class, 'ability' and attainment: a problem for the sociology of education. British Journal of Sociology of Education, Vol.22 (2), pp. 189-202.

Ocampo, J. (2000). *Equidad, Desarrollo y Ciudadanía*. CEPAL Documento Central 2000-2001. Santiago: CEPAL.

Opdenakker, M. y Damme J. (2000). Effects of schools, teaching staff and classes ion achievement and well-being in secondary education: similarities and differences between school outcomes. *School Effectiveness and School Improvement*, Vol.11 (2), pp. 165-196.

Perrenoud, Ph. (1990). *La construcción del éxito y del fracaso escolar*. Madrid: Morata - Paideia.

Ridell, A. (1993).The evidence on public/private educational trade-offs in developing countries. *International Journal of Educational Development*, 13, pp. 373-386.

Ridell, A. (1997). Assessing designs school effectiveness research and school improvement in developing countries. *Comparative Education Review*, vol.41 (2), pp. 178-204.

Raudenbush S. y Bryk, A. (1986). A hierarchical model for studying school effects, *Sociology of Education*, Vol. 59 (January), pp. 1-17.

Raudenbush, S. y Willms J. (1995). The estimation of school effects", *Journal of Educational and Behavioral Statistics*, .20 (4), pp. 307-335

SIEMPRO (2001). *No se trata de un ranking de calidad, se trata de un ranking de desigualdad y pobreza*. Sistema de

Información, Monitoreo y Evaluación de Programas Sociales - SIEMPRO. Buenos Aires: SIEMPRO.

Tedesco, J. (1992) Privatización educativa y calidad de la educación en América Latina. *Propuesta Educativa*, 6, pp. 5-14.

Teddlie, C. y Reynolds, D. (2001). Countering the critics: Responses to recent criticism of school effectiveness research. *School Effectiveness and School Improvement*, Vol.12, Nº 1: 41-82

Thrupp, M. (2001a). Sociological and political concerns about school effectiveness research: Time for a new research agenda. *School Effectiveness and Scholl Improvement*, Vol. 12 (1), pp.7-40.

Thrupp, M. (2001b) Recent school effectiveness counter-critiques: problems and possibilities. British Educational Research Journal, Vol. 27, Nº4, pp. 443-457.

Zibas, D. (1996) La agenda latinoamericana de modernización educativa y la privatización de la enseñanza media. *Propuesta Educativa*, 14, pp. 32-39.

Acerca del Autoro

Rubén Alberto Cervini
Universidad Nacional de Quilmes, Argentina
Dirección: Salguero 2460 5º Piso Capital Federal c.p. 2514 - Argentina
Dirección electrónica: racervin @infovia.com.ar

Profesor titular e investigador en el Departamento de Ciencia Sociales de la Universidad Nacional de Quilmes, a cargo de los cursos "Metodología de la Investigación Educativa" y "Evaluación de Programas y Sistemas Educativos" de la carrera de Educación. Investigador responsable del proyecto "La Educación Media en la Argentina". Publicaciones recientes:

- 1999 "Los factores del rendimiento de Matemática en la escuela primaria urbana de Argentina" En: Memoria del Seminario Iberoamericano - 1997, OEI/Ministerio de Cultura y Educación de Argentina.
- 1999 "Calidad y Equidad en la Educación Básica de Argentina", Serie: *Factores Asociados al Logro Escolar*, nº 5, Ministerio de Cultura y Educación de la Nación, Buenos Aires.
- 1999 "Factores escolares del rendimiento en Matemática y Lengua 7º año de la EGB". En: *Los factores Escolares Asociados a Logros Académicos - Segunda Investigación en Mendoza*. Sistema Público de Evaluación de la Calidad, Gobierno de Mendoza.
- 2001 "Efecto de la 'Oportunidad de Aprender' sobre el Logro en Matemáticas de la Educación Básica de Argentina. *Revista Electrónica de Investigación Educativa REDIE*. Vol. 3, Nº 21. México: Universidad Autónoma de Baja California. (http://redie.ens.uabc.mx)
- 2002 "La distribución social de los rendimientos escolares". En: *El rendimiento escolar en Argentina - Análisis de resultados y factores*. Emilio Tenti (Org.), Cap.IV. Buenos Aires: Editorial Losada.
- 2002 "Los factores del rendimiento en la educación básica de Argentina." En: *El rendimiento escolar en Argentina - Análisis de resultados y factores*. Emilio Tenti (Org.), Cap. V. Buenos Aires: Editorial Losada.
- 2002 "Desigualdades en el Logro Académico y Reproducción Cultural en la Educación Primaria de Argentina - Un modelo de tres niveles." *Revista Mexicana de Investigación Educativa*, Vol. 16, 2002. México: Consejo Mexicano de Investigación Educativa. (dictaminado para publicación)
- *Participación familiar y logro académico del alumno*. Revista Colombiana de Educación. Colombia: Universidad Autónoma Nacional de Colombia. (Dictaminado para publicación segundo semestre de 2002).
- *Desigualdades Socioculturales en el Aprendizaje de Matemática y Lengua de la Educación Secundaria en Argentina - Un modelo de tres niveles. Revista Electrónica de Investigación y Evaluación Educativa* Universidad de Valencia, España. (Dictaminado para publicación en el actual "Número en construcción")

Copyright 2002 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

This article has been retrieved 3347 times since February 16, 2003

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Education Policy Analysis Archives
Volume 11 Number 7 **February 16, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the **EDUCATION POLICY ANALYSIS ARCHIVES**. Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Impact of Minnesota's "Profile of Learning" on Teaching and Learning in English and Social Studies Classrooms

Patricia G. Avery
Richard Beach
Jodiann Coler
University of Minnesota

Citation: Avery, P. G., Beach, R., & Coler, J. (2003, February 16). The impact of Minnesota's "Profile of Learning" on teaching and learning in English and social studies classrooms, *Education Policy Analysis Archives, 11*(7). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n7/.

Abstract

In 1990, the Minnesota State Board of Education declared its intention to develop a "results-oriented graduation requirement" based on student achievement as opposed to the usual credit/course completion requirement. In addition to a traditional test of basic skills, the state began developing the Profile of Learning, a set of performance-based standards grounded in a constructivist educational philosophy, an approach that differs from the content-based standards found in many states. The Profile was controversial from its inception. Conservatives characterized the Profile as too process- oriented and as lacking subject-matter content; teachers reported that the Profile required a significant amount of additional teacher preparation time; and parents, who were not adequately informed about the Profile, questioned the purpose of the Profile. Teachers were frustrated with the confusing and sometimes contradictory directions they received from the Minnesota Department of Children, Families, and Learning charged with implementing the Profile. In 2000-2001, we surveyed and interviewed selected secondary English and social studies teachers in the state about their perceptions of the Profile's impact on teaching and learning. Among the positive perceptions was an increase in students' higher order thinking, students' understanding of criteria for quality work, and teachers conversations with one another about instructional issues. Increased teacher preparation time and decreased enjoyment of teaching were among the negative perceptions. Teachers also experienced difficulty adopting performance assessment techniques. When teachers believed they received effective preparation and adequate resources for working with the Profile, they were much more likely to report beneficial effects in terms of teaching and learning. The majority of teachers, however, rated their preparation and resources as "fair" or "poor." Results are discussed in terms of school and instructional change.

At the beginning of the year 2000, 49 of 50 states had adopted standards that describe what students should "know and be able to do." Many of the standards documents were created to set high academic expectations for all students, and to add "rigor" to purportedly watered-down curricula. Most states have developed or are developing assessments to determine whether students "meet the standards"—hence the term "standards-based assessments." Thirty-seven of the states' assessments reflect yet another recent development in education—a trend toward the use of nontraditional assessments (*Education Week*, 2000). The nontraditional assessments range from constructed response items (short answer) to demonstrations of performance, such as conducting a science experiment or giving a persuasive speech. In contrast to the use of multiple-choice tests, the use of performance assessments is thought to challenge students in ways that allow for individual strengths and diversity in thinking (Eisner, 1999; Wiggins, 1998).

We are thus witnessing two major changes in education: standards-based assessment and performance-based assessment—both of which are being conducted in many states for high stakes. Evaluation and research studies on the implementation and effects of state standards-based assessments are just now beginning to accumulate. Some of the state standards documents are perceived by teachers and the public as confusing and overly burdensome (McDonnell & Choisser, 1997; Schomaker & Marzano, 1999). Many reports suggest that teachers are ill prepared to use nontraditional assessments (Bateson, 1994; Black & Wiliam, 1998; Firestone, Roseblum, & Bader, 1992; Plake & Impara, 1997). When teachers *are* comfortable with nontraditional assessments, studies find that the format requires an enormous amount of teacher time in addition to the costs of scoring (Koretz, Stecher, Klein, & McCaffrey, 1994). Some states, including Arizona, California, Kentucky, and most recently Maryland, have pulled back from the idea of using high stakes performance-based assessments because of concerns about time, cost, and questionable psychometric properties. A few studies report teachers perceive positive changes in their instruction when they use performance assessments, but the same teachers question whether the costs make the change worthwhile (Herman, 1997; Koretz et al., 1994; Madaus & Kellaghan, 1993).

Underlying the development of performance assessments is a constructivist philosophy toward teaching and learning. Although there are various interpretations of constructivism among scholars, most agree that it implies that students "construct" meaning by engaging in activities that require them to manipulate and synthesize data, rather than reproduce information. Teachers in states adopting constructivist-oriented standards often have difficulty switching to authentic or performance-based assessment of students' demonstration of learning. Missouri teachers experienced considerable difficulty implementing performance assessments, mostly due to lack of training (Jackson, 2000). An analysis of high school teachers in three suburban Illinois schools indicated that only a small number of these teachers were actually using authentic assessments (Meisenheimer, 1996). Those teachers who did employ performance assessments were more likely to be receiving in-service training, were actively involved in professional organizations and in their schools, and had a strong philosophical understanding of the purpose and value of authentic assessment. It was also the case that these teachers were working in schools that supported their efforts by encouraging their experimentation and providing them with in-service training.

In this report, we describe the state assessment system developed in Minnesota, a state that has long had a reputation for being innovative and progressive in the area of education. We briefly relate significant events in the "story" of the Minnesota standards, and then present the quantitative results of a survey of English and social studies teachers on their perceptions of the impact of the standards. We also identify themes and issues that emerge from the qualitative survey data, as well as interviews conducted with selected teachers. Together, survey and interview data provide insights into the promise and challenge of standards-based reform, particularly as it relates to constructivist-oriented, performance-based assessment.

The Development of Minnesota's Graduation Requirements

The current national focus on standards-based assessment is often traced back to the publication of *A Nation at Risk* in 1983, a report from then Secretary of Education Terrell Bell (National Commission on Excellence in Education, 1983). The report characterized the public education system as a "rising tide of mediocrity" that no longer prepares young people for adult work and responsibilities. According to the report, watered-down content, low expectations for students, and poorly prepared teachers had contributed to a weak and deteriorating educational system. *A Nation at Risk* prompted a wave of educational reports from national and state commissions, "Blue Ribbon" panels, and community leaders, each designed to give us a picture of the "status of education," either in a particular area (e.g., teacher education programs, middle schools), for a specific group (e.g., low-income students, special needs students), or for a certain locale (e.g., state, region).

Minnesota business and community leaders, concerned that too many high school graduates did not have basic math and literacy skills, joined the call for educational reforms that would better prepare young people for the workforce. The notion that "seat time" should not qualify students for a high school diploma shifted attention toward "what students know and can do" as the criteria for graduation. In 1987, the Minnesota legislature directed the State Board of Education to identify "core learner outcomes" for each curriculum area, i.e., what should students know and be able to do in mathematics? in social studies? in English? The first set of Essential Learner Outcomes was adopted by the State Board of Education in 1988.

At the national level in 1989, President George H. Bush convened an education summit with the nation's governors in Virginia. The group agreed on six education goals to be achieved by the year 2000; these goals were collectively referred to as "America 2000." President Bill Clinton later added two goals, and renamed the list "Goals 2000." The first goal, and the one most often cited, states that "All children will start school ready to learn." It is the third goal, however, that bears directly on the standards-based assessment movement:

> American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter including English, mathematics, science, history and geography. (*America 2000,* 1991, p. 9)

This particular goal is consistent with the trajectory Minnesota followed in terms of educational reform.

In 1990 the Minnesota State Board of Education declared its intention to develop a "results-oriented graduation requirement" based on student achievement as opposed to the current credit/course completion requirement. A Graduation Standards Executive Committee, composed of business, education and citizen groups, was appointed to review the process of moving toward this "results-oriented" system. It was about this time that a group of education scholars conducted an in-depth study of the assessment reforms underway in six states, among them Minnesota. Their observations were published in the *Teachers College Record* in 1992 (Firestone et al., 1992). The authors suggested that Minnesota's plans for reform (along with Arizona's plans, the ASAP test) were notable because they held the potential to increase students' higher-level thinking.

The Minnesota Department of Children, Families and Learning (CFL) sought to develop two sets of standards, one focusing on basic skills in math, reading and writing, and another designed to set high, rigorous expectations for students. The former would be assessed through the Minnesota Basic Skills Test, and the latter through performance-based assessments called the *Minnesota High Standards* (formerly called, and most commonly still referred to as the *Profile of Learning*). Both the basic and high standards are purportedly guided by five Comprehensive Goals—that students who graduate from the Minnesota public schools be:

- Purposeful Thinkers;
- Effective Communicators;
- Self-Directed Learners;
- Productive Group Participants; and
- Responsible Citizens.

The Basic Skills Test. In order to graduate, all public school students in Minnesota, beginning with the high school graduating class of 2000, were required to pass the Basic Skills Tests in reading and math. These tests are written in the traditional multiple-choice format. The Basic Skills tests are initially given to students in the eighth grade, and students who do not pass the test can take it annually through the twelfth grade. The Basic Skills Test in writing composition is given in the tenth grade, and similar to the reading and math tests, students who fail the writing test can re-take it through the twelfth grade. The class of 2001 was required to pass the writing test (in addition to the reading and math tests) as a condition for graduation. All of the basic skills tests are "high stakes tests"—students who do not pass these tests are not to receive a high school diploma.

The basic skills tests generated little controversy in Minnesota until the summer of 2000, when it was discovered that a data entry error had incorrectly scored one form of the math basic skills test. Approximately 8,000 students were told they had failed the test, when in fact they had passed the test. Of these students, approximately 300 were seniors who were not permitted to graduate with their class in the spring (Welsh, 2001).

Special legislative sessions were convened during the summer to determine how such an error could have occurred. The situation prompted many Minnesotans to question the wisdom of using the score from one test to determine whether a student should graduate. At present, however, the major change in the system has been the implementation of a range of safeguards to lessen the likelihood that such an error will occur again.

The Profile of Learning. The *Profile of Learning* has generated the most controversy in the state of Minnesota. Not surprisingly, it also represents a significant deviation from traditional schooling and testing. Whereas the Basic Skills Tests set a minimum level of knowledge for students to attain, the *Profile* required students to demonstrate a higher level of understanding through performance-based assessments. Similarly, while the Basic Skills Tests focuses on traditional subject areas (reading, writing and math), the *Profile* was originally based on interdisciplinary "learning areas" that characterize a "well-rounded" education. The 110 Essential Learner Outcomes identified by CFL in the early 1990s were reworked to form a list of 25, then 15, and finally, 10 areas of learning. Table 1 shows the changes in the learning areas between 1993 and 2002. The learning areas that were most often used between 1993 and 2000 reflected an effort to move toward major interdisciplinary concepts (e.g., people and cultures, decision making) and toward more active, practical learning (e.g., mathematical *applications*). The current learning areas include the names of more traditional subject areas, such as social studies, physical education and economics.

Table 1
Minnesota's Learning Areas: 1993 – 2002

Learning Areas 1993-2000	Learning Areas 2001-2002
Read, View and Listen	Read, Listen and View
Write and Speak	Write and Speak
Literature and the Arts	Arts and Literature
Mathematical Applications	Mathematical Concepts and Applications
Inquiry	Inquiry and Research
Scientific Applications	Scientific Concepts and Applications
People and Cultures	Social Studies
Decision Making	Physical Education and Lifetime Fitness
Resource Management	Economics and Business

World Languages (optional)	World Languages (optional)

Note: There have been many changes in the Learning Areas in the past decade; however, these were the dominant areas for the time periods shown.

Each learning area encompasses two or more content standards. For example, in 2000, a middle grades content standard associated with the "Read, View and Listen" learning area was as follows:

> *Literature and Arts Analysis and Interpretation.*
> A student shall demonstrate the ability to interpret and evaluate complex works of music, dance, theater, visual arts, literature, or media arts by doing the following:
>
> A. describing the elements and structure of the art form; the artistic intent; and the historical, cultural, and social background of the selected art works;
> B. applying specific critical criteria to interpret and analyze the selected art works;
> C. describing how particular effects are produced by the artist's use of the elements of the art form; and
> D. communicating an informed interpretation using the vocabulary of the art form.

A high school content standard often associated with the social studies under the learning area "Inquiry" was as follows:

> *Issue Analysis.*
> A student shall research an issue and evaluate proposed positions or solutions by:
>
> A. gathering information on past or contemporary issues;
> B. identifying relevant questions or a range of points of view;
> C. summarizing relevant background information;
> D. examining information from each source for bias and intended audience;
> E. identifying areas of conflict, compromise, or agreement among various groups concerning the issue; and
> F. evaluating multiple positions and proposed solutions for the issue, including analyzing conclusions, arguments, and supporting evidence; identifying motives of groups or individuals; analyzing feasibility and practicality; identifying impact on policies; comparing alternative solutions; and projecting consequences.

In order to graduate, students were to complete 24 standards in grades 9 – 12. Students would receive credit for attempting a standard, even if their work was unsatisfactory. Scores were to be recorded on student transcripts. Students in grades 1-8 were to complete "preparatory standards," the "building blocks" for the high school standards. In essence, the *Profile* created a "spiral performance assessment system" around 10 major themes; the performances became increasingly complex within a given theme or "area of learning" as students progressed through school (Minnesota Department of Children, Families, and Learning, 1998).

To foster implementation of the *Profile*, CFL relied primarily on a "train-the-trainer" model. Training workshops on the *Profile* were organized throughout the state for selected teachers and administrators. These representatives then returned to their own districts to provide teachers with training at the local level. Districts also designated certain curriculum coordinators as responsible for overseeing the *Profile* implementation. Teachers met in local districts to discuss ways of aligning their own curriculums to the standards, meetings that sometimes involved extensive rethinking of their teaching.

One problem with relying primarily on a "train-the-trainer" model was that, other than basic information on the CFL Website and the packages, there were few alternative sources of information—printed materials, videos, or curriculum frameworks for teachers, administrators, and parents. (One rationale for the lack of print materials was that because the *Profile* was continually changing, CFL was reluctant to print materials that would become outdated or outmoded. Materials and handouts from CFL rarely indicated dates or authors. It was therefore difficult to ascertain whether particular policies had been superceded by other policies, adding to teacher confusion over policy).

This "train-the-trainer" model effectively served those teachers who were willing to attend workshops and actively participate in the training process, especially in districts that were providing high levels of support. However, a sizable number of teachers who were less enthusiastic about the *Profile* often received only minimal training from individuals who, through no fault of their own, were not always familiar with the most recent changes in the *Profile*. At workshops it was not unusual to have teachers sharing conflicting information they had received from people who should have been "in the know." As a result, these reluctant teachers, as well as parents and the public, often had little understanding of the *Profile*.

In workshops across the state, teachers were told that students should demonstrate they had "met the standard" through high-quality "performance packages." A "performance package" is defined as a set of interrelated performance tasks that give students the opportunity to demonstrate mastery of a standard. The "performance packages" were to be "embedded" into the curriculum.

A CFL handout entitled "The A, B, C's of Performance Tasks," stipulated that the performance tasks in the packages should be authentic, unbiased, and constructivist. CFL developed "performance packages" to serve as models for teachers, and eventually, most of the standards were accompanied by "performance packages." In the "performance package" designed to meet the "Issues Analysis" standard previously cited, for example, students were required to research an issue of importance to them, identify key stakeholders and interest groups related to the issue, prepare a position paper stating their own beliefs about the issue, develop a consensus position among a small group of peers, and present their findings to a community group involved in the decision-making processes that affect the issue.

Many aspects of the "performance packages" were consistent with the characteristics of "effective instruction," as well as major principles of learning and motivation. For example, at various points in most of the "performance packages" were checklists of tasks the students were to complete. The checklists, in addition to specifying the criteria by which the work would eventually be evaluated, required students to self-assess, and teachers to monitor students' progress. The checklists assured that students would receive feedback throughout their work.

The students were often required to be active participants in "constructing their own meaning" by collecting or manipulating data, posing hypotheses and making generalizations. Successful completion of a "performance package" frequently required students to work in cooperative groups, or to interact with community members outside the classroom. Theodore Sizer might call many of the students' work products "exhibitions" (Sizer, 1997); the *Teaching for Understanding* group at Harvard University might call the students' work "performances of understanding" (Wiske, 1998); and Fred Newmann and his colleagues of the former Center on Organization and Restructuring of Schools (CORS) at the University of Wisconsin might label the students' work examples of "authentic student performance" (Newmann, Secada, & Wehlage, 1995). Although these scholars would undoubtedly make changes in the "performance packages," they would probably be supportive of the philosophy upon which the packages were based.

When students completed a "performance package," their work associated with meeting a particular standard was evaluated by their teacher on criteria specified by a state rubric, and awarded a holistic score of 4 (exemplary), 3 (proficient), 2 (novice) or 1 (beginning). Students could meet five of the six criteria listed under "4", but if the sixth criteria merited a "2", students would be awarded a "2." In other words, all parts of the listed criteria needed to be met for a specific score to be given.

Although many teachers found some merit in specific aspects of the "performance packages," the packages became a focal point for a barrage of criticism from teachers, parents, and community members. Schomaker and Marzano (1999) note that "most of the state assessment-based standards documents have contributed to the problem they were designed to address. Documents are way too long, and full of educational jargon." Unfortunately, the *Profile,* and more specifically its accompanying performance packages, are subject to their critique. The packages used terminology unfamiliar even to veteran teachers (e.g., "content standard," "element," "task management skills"). They required teachers to use skills with which many were unfamiliar, such as using checklists or scoring rubrics. Some of the performance packages required content knowledge that teachers simply had not acquired. The sheer length of the packages (one was 65 pages!) was overwhelming to students and teachers alike. Many teachers complained that the performance packages were becoming the *de facto* curriculum. Moreover, the quality of the performance packages developed by the state was uneven.

Although the CFL developed the performance packages to serve as *models,* many districts either assumed the packages were state-mandated, or because the development of a package was so time intensive, mandated the use of the state packages within their district. And while CFL insisted that districts could develop their own performance packages, the Department also wanted some kind of "quality control" to assure that all students were expected to demonstrate the same level of academic rigor. Initially, CFL wanted to monitor the quality of the performance packages. Then, because of vehement cries for local control, individual school districts gained the authority to give their "stamp of approval" to performance packages developed within the district. Most recently, CFL declared that separate performance *assessments,* instead of performance *packages,* can be used to meet parts of a standard. The significance of this is that while the completion of a performance package was often burdensome and overwhelming for a one-quarter civics class, for example, completion of short performance assessments could more easily be interwoven into an existing course.

Aside from their concerns about the performance packages, teachers grew frustrated with the constant changes in the standards requirements. Similarly, it was not unusual to get contradictory information from CFL representatives. CFL wanted to be attentive to teachers' feedback about the *Profile,* but in doing so, this often meant making changes that further frustrated teachers.

In 1993, the Minnesota State Legislature envisioned that both the Basic Skills Tests and the *Profile of Learning* would be required of students entering ninth grade during the 1996-97 school year. The Basic Skills Tests in reading and mathematics were in place for ninth graders in 1996-97 (the class scheduled to graduate in 2000); the Basic Skills Writing Test was deferred until the following year. Beginning with the ninth grade class of 1998, students were to have completed the *Profile* in order to graduate.

The *Profile of Learning* was the subject of intense debate in the 1998, 1999 and 2000 legislative sessions. In 1998, the legislature created a Standards Advisory Panel, composed of 11 leaders from business and education, to make recommendations to the 1999 legislature on the implementation of the *Profile of Learning*. Among their recommendations, the Advisory Panel suggested that the number of "learning areas" be reduced from 10 to 5; that the reference to state performance packages in the Graduation Rule be eliminated, and that the language used in the Graduation Rule be "clear and understandable to teachers, parents and students."

During the 1999 legislative session, the House voted to eliminate the *Profile* in favor of more traditional coursework. The Senate voted to retain the *Profile,* but with some of the modifications suggested by the Standards Advisory Panel. The session ended without any action taken on the *Profile*. House conferees refused to consider modifications to the *Profile;* had the Legislature adopted the modifications, it was thought that the widespread opposition to the *Profile* would have decreased substantially. The goal of the staunch opponents to the *Profile* was to eliminate it, not to modify it.

In early 2000, a poll released by the state teachers' union, Education Minnesota, indicated that 39% of the 608 teachers surveyed wanted to eliminate the *Profile* altogether; another 51% wanted significant changes; *and only 9% of the teachers believed the Profile should remain in its current form* (Draper, 2000). Education Minnesota co-presidents called for a major overhaul of the *Profile*. At about the same time, *Education Week* published a report entitled *Quality Counts* in which they graded states on their assessment programs. States received a grade based on the types of assessments used, and the number of subject areas assessed. Minnesota, ranked in the bottom 10 states, was given a grade of C- (*Education Week,* 2000).

In the 2000 legislative session, the *Profile* narrowly escaped elimination. The House of Representatives voted 97 – 34 to delay indefinitely the implementation of the *Profile* as a graduation requirement. Conservatives proposed the North Star Standard, a plan that focuses on the "basics" in core subject area courses, as an alternative to the *Profile*. The North Star Standard would focus on content over process, and would use the traditional A-F grading system as opposed to the 4-3-2-1 scores mandated by the *Profile*.

The Senate was generally more supportive of the *Profile,* and in a conference committee convened in May 2000, a compromise was reached whereby local school boards would be allowed to choose between the *Profile* and the North Star Standard, the back-to-basics alternative inspired by the House. *Profile* supporters believed that the only bill that would pass both the House and the Senate needed to include the *Profile* and the North Star Standard. However, *Profile* supporters, together with the CFL Commissioner, insisted that students in both *Profile* and North Star Standard districts take the Minnesota Comprehensive Assessments, standardized tests for school accountability based on *Profile*-related goals. At the last minute, the North Star Standards supporters refused to sign the compromise bill, ostensibly because it required assessments that did not match the goals of their back-to-basics standards. The Senate passed a "*Profile*-only" bill, 82 – 44 at 3:20 a.m. on May 18th. An hour and one-half later, members of the House cast the last vote of the longest legislative session in Minnesota history, and passed the "*Profile*-only" bill 99-27.

The bill approved by the House and Senate gave districts much more control over the way in which the *Profile* was (or was not) to be implemented. Each district's teachers, administrators and school board members were to vote on how many, if any, standards from the *Profile* students would need to complete. The bill encourages districts to work toward implementation of all 24 standards, but no timeline is mentioned. Slightly over half (53%) of the state's 332 districts voted to require all 24 standards. A few districts voted not to require their students to complete any standards. Local districts could decide whether to use the familiar letter grades as opposed to the 4-3-2-1 system. Significantly, the statute also stated that "districts...may use one or more assessment methods to measure students' performance on one or more content standards. The commissioner [of the Department of Children, Families and Learning] shall not mandate in rule or otherwise the assessment methods that local sites must use to meet the requirements under this section."

Opponents of the *Profile* vowed to renew the fight to eliminate the *Profile* in the 2001 legislative session. Lawmakers, however, seemed weary of the *Profile* debate in 2001. School funding formulas and early childhood education were the focus of attention in terms of educational issues. A state budget crisis dominates the legislative agenda in 2002. Until April 19, 2002, there had been little discussion of the *Profile*. But on that date, the House Majority Leader introduced an amendment to repeal the *Profile*. The amendment won bi-partisan support, and passed 109 – 22 (Bakst, 2002). The vote in the Senate was tied, 33 – 33. Although Governor Jesse Ventura had supported the *Profile,* and thus would most likely veto a proposal to eliminate it, the "near-death" experience of the *Profile* jarred many of its supporters (Lonetree, 2002). In the annual *Quality Counts* report published in *Education Week* in early 2002, Minnesota's "grade" for "Standards and Accountability" dropped to a D- (*Education Week,* 2002). Tim Pawlenty, a Republican, was elected to serve as Minnesota's governor in the November 2002 election; a major part of his campaign platform was a promise to eliminate the *Profile*. Thus, at the time of this writing, the future of the *Profile* is tenuous at best.

Over the past five years, several studies have examined the implementation of the *Profile*. A 1998 survey administered to a sample of 1600 teachers from 100 Minnesota public schools asked teachers to assess their knowledge and understanding of the *Profile*. Over 80% of the teachers indicated they knew enough about the *Profile* to integrate the standards into their teaching (Human Capital Research Corporation, 1998). A later study based on focus groups with teachers across the state examined the degree to which the standards are being implemented in the schools (Minnesota Department of Children, Families & Learning, 2000). Almost two-thirds of the 2500 teachers who participated in the focus groups believed the standards had been integrated into their curriculum, but only half felt the standards were "aligned" with instruction, assessment and curriculum. One of the more significant themes of the focus groups was that teachers believed they were talking with one another more often about curricular issues.

Notably absent have been studies of how the *Profile* is affecting school and classroom practice. In the present study, we begin to shift the focus of research and evaluation away from teacher knowledge and implementation issues, and toward the impact of the *Profile* on teaching and learning in the classroom. We begin by asking teachers their *perceptions of* the *Profile's* impact on teaching and learning.

Methods

The respondents. All 292 English/language arts teachers who are members of the Minnesota Council of Teachers of English (MCTE) were surveyed; 171 or 59% returned completed questionnaires. Similarly, the 945 secondary social studies teachers who are members of the Minnesota Council for the Social Studies (MCSS) were surveyed; 487 or 52% of the teachers completed the questionnaire. Table 2 provides demographic information about the teachers. The typical respondent was a mid-career European-American with a Masters degree who taught outside the inner city. Follow-up interviews were conducted with 51 English teachers and 89 social studies teachers.

Table 2
Characteristics of Survey Respondents and their Schools
(N = 658)

Gender	
Female	50%
Male	50%
Highest Degree	
Bachelor of Arts	37%
Masters	59%
Specialist Certificate	3%
Other	1%
School Setting	
Rural	45%
Suburban	44%
Urban	11%
Length of Class Periods	
30-55 minutes	71%
Over 55 minutes	29%
Years of Teaching Experience	
Range:	0-39 years
Mean:	17 years
Size of Social Studies Department	
Range:	1-25 teachers
Mean:	8 teachers

The questionnaire. A questionnaire was designed to assess secondary English and social studies teachers' perceptions of the ways in with the *Profile* may be affecting teaching and learning in their classrooms. Eight items focus on how the *Profile* may have impacted student learning (e.g., students' higher level thinking, students' interest in social studies) and 10 items focus on how the *Profile* may have affected teaching (e.g., teachers' preparation time, teachers' enjoyment of teaching). For each item, teachers were asked to respond on a 7-point scale (decreased a lot, decreased moderately, decreased slightly, no impact, increased slightly, increased moderately, increased a lot). Items were primarily chosen because they represent (1) goals associated with the study of English and social studies (e.g., students' interest in English/social studies); (2) characteristics of "authentic pedagogy"(Note 1) (e.g., students' interactions with one another about social studies content); (3) characteristics of professional community (Note 2) (e.g., teachers' conversations with colleagues in their school about English/social studies instruction and assessment); and (4) stated goals of CFL (e.g., the degree to which students are prepared for "life after graduation"). These categories are not mutually exclusive. For example, increasing students' higher level thinking is not only a goal of the English/social studies and CFL, but is also a characteristic of authentic pedagogy. Finally, standard demographic information was collected (e.g., years of teaching experience, school setting, most advanced degree), and teachers were asked about the quality of their preparation for working with the *Profile* (e.g., teacher/inservice workshops).

The first draft of the questionnaire was reviewed by two expert social studies teachers, both of whom have earned National Board Certification, one curriculum coordinator for an urban district, a Minnesota state curriculum coordinator, and an English education professor with expertise in the state standards movement. The questionnaire was revised several times based on reviewers' comments and suggestions.

The interviews. Interviews were conducted with a selected number of teachers completing the questionnaire. Teachers were asked to elaborate on their written comments on the questionnaire (regarding positive and negative aspects of the *Profile)*, and to describe their work with standards packages/assignments. In some cases, we deviated from the interview schedule to ask follow-up questions. Transcripts of interviews were analyzed in terms of references to various topics.

Data collection procedures. Questionnaires were mailed the second week in September 2000 to all MCTE and MCSS secondary teachers. A postcard reminding teachers to return the questionnaire was mailed one week later, and two weeks after the postcard mailing, a second copy of the questionnaire was mailed to teachers who had not yet responded to the survey. On the questionnaire, teachers indicated whether they would be willing to be interviewed about their responses. Phone or e-mail interviews were conducted in February and March of 2001 with those teachers who indicated a willingness to be interviewed.

Data analysis. For the questionnaire items, a simple frequency of responses was calculated. NVIVO, a qualitative data analysis software package, was used to code the interview data. All three researchers initially read 25% of the transcripts to create a set of closed codes to use in the analysis of the interview data. Through an iterative process, agreement was reached on coding categories. Each investigator then coded one-third of the data, after which the coded data were then reviewed for emergent themes and frequency of responses in coding categories.

Results

In this section, we report the results of the quantitative analysis, and offer excerpts from the interview data to explain some of the quantitative findings. Thus, we weave together questionnaire and interview data to present a picture of how teachers perceive the *Profile* is impacting teaching and learning in their classrooms. The interview data also suggested teacher observations and concerns not directly related to the questionnaire items. In an effort to give voice to these teachers, the themes and patterns in these data are also described.

Impact on student learning. Table 3 shows teachers' perceptions of the impact of the *Profile* on students' learning. The percentage of teachers who believe the *Profile* has had a positive impact on student learning ranges from 22% (increased student interest in English/social studies) to 51% (increased students' higher level thinking). In many, though not a majority of classrooms, teachers perceive the *Profile* to be having a positive impact on student learning.

Table 3
Respondents' Perceptions of the Impact of the
***Profile of Learning* on Student Learning**
(N = 658)

	Decreased (%)	**No Impact (%)**	**Increased (%)**
Students' higher level thinking	6%	43%	51%
Students' interest in English/social studies	28	50	22
Students' interaction with one another	9	54	37
Students' understanding of grading criteria*	22	33	45
Teacher communication with students about work quality	7	46	47
Students' interaction with community outside school	4	65	31
Quality of students' work on assignment	11	55	34
Students' preparation for "life" after school	9	62	29

*English teachers were significantly more likely to perceive increases in students' understanding of grading criteria than were social studies teachers.

In the interviews, teachers noted that constructivist instruction requires students to take responsibility for their own learning, apply their own knowledge, and work together collaboratively. They also noted the value of having students demonstrate proficiency through "hands-on" learning associated with higher levels of student involvement in the classroom. One teacher cited a specific example from a class simulation of the Treaty of Versailles she created to meet a content standard:

> I can remember one young man two years ago who got into being part of the Turkish delegation at the Treaty of Versailles, and came up with original pieces of documentation that he just loved. And it turned him onto history. That's the payback. When you see the light bulb turn on and history becomes more than a textbook or a dry set of facts.

For teachers such as this one, the *Profile* offered an opportunity to set high expectations for students and to assess their learning in a constructivist manner.

Some teachers noted that the *Profile's* focus on authentic, "hands-on" assessment was already consistent with their own previous constructivist instruction. One teacher noted that the *Profile* is "fine because I've done a lot of hands-on activities in my classroom. And I believe in that—show me what you learned, not just tell me on a sheet of paper." Teachers who already used constructivist strategies in their classrooms agreed with the theory behind the *Profile*, yet also saw it as redundant for their instructional practices.

Impact on teaching. Table 4 documents teachers' perceptions of the impact of the *Profile* on aspects of their teaching. Similar to teachers' perceptions of the impact of the *Profile* on student learning, many teachers perceive the *Profile* to be having a positive impact on various aspects of their teaching. Slightly more than one-third believe the *Profile* has helped to increase the coordination of content across grade levels, the range of teachers' instructional methods and materials (among them computer technology), and the use of nontraditional assessments. Teachers reported that rather than focus on their own instruction, they had to focus on student learning because the *Profile* required them to explain their learning expectations to students and parents, clarify criteria for evaluation, share these criteria with colleagues, and display student work. Most educational reformers would view these developments as potentially positive.

Table 4
Teachers' Perceptions of the Impact of the
***Profile of Learning* on Teaching**
(N = 658)

Item	**Decreased**	**No Impact**	**Increased**

	(%)	(%)	(%)
Teachers' preparation time for classes	7%	11%	82%
Use of a wider range of teaching materials	9	44	47
Use of nontraditional assessments	10	55	35
Conversations with school colleagues about social studies teaching and assessment	3	31	66
Interaction with colleagues outside school	3	56	41
Interest in subject area	17	67	16
Enjoyment of teaching*	53	35	12
Coordination across grade levels	10	51	39
Use of computer technology	5	61	34
Use of different teaching approaches	6	58	36

*Social studies teachers were significantly more likely to indicate a decrease in their enjoyment of teaching than were English teachers.

Perhaps most striking is the teachers' report that the *Profile* prompted more discussion about English/social studies instruction and assessment with their colleagues. As one teacher noted, teachers were more likely to discuss issues of curriculum development given the mandate of standards implementation:

> We have talked more, had more opportunities to connect....the *Profile* is a good "equalizer" for staff in various curricular areas—more understanding/integrating. Even staff who have disliked/discounted standards/profiles have a greater sense of purpose as professionals.

In reflecting on both local and state-wide training, the opportunity to have positive professional discussions with colleagues was seen as a positive outcome.

The school culture has traditionally isolated most teachers from one another in terms of substantive conversations about their work. Although many of their conversations might have been based on complaints about the *Profile*, it is quite likely that the discussions increased teachers' sense of collegiality, as well as their understanding of one another's views on high quality instruction and assessment.

Although the *Profile* appears to have had a positive impact in many classrooms, teachers perceive at least two very strong negative aspects to the *Profile:* More than four-fifths of the teachers believe the *Profile* has increased their preparation time, and over one-half report that working with the *Profile* has decreased their enjoyment of teaching.

By far the most frequently mentioned issue for teachers in regards to implementation in our interviews was the "huge amount of time invested." More time was spent in the preparation of the packages, pre-teaching in class, completing the performance assessments in class, grading the performance assessments, and record-keeping and documentation.

In the interviews, teachers reported that conscientious teachers devoted considerable time to learning how to the implement the standards:

> The pressure to prove to someone that they were doing a good job put unnecessary pressure on those teachers, took time away from other areas to go to training sessions and create whatever for their districts...just additional time away from other classroom activities or work days when they were planning or preparing for the next quarter.

In many cases not only did teachers see the additional time as an issue, but also felt that the loss of time may have been equally well spent on other areas of curriculum, instruction, and assessment.

Rather than relying solely on textbooks, teachers often had to devise their own curriculum materials, frequently without adequate financial support for such materials. One teacher described the preparation involved:

> I've done the time capsule, where they, students are supposed to select 10 items or 10 events or people that really influenced the United States from 1900 to 2000. And then they're supposed to come up with ideas on [that item.] They write a description of it, and then they write another paragraph justifying it. And getting all the materials ready, getting the library time to do that, when they're so limited. I mean, the state asked us to do this, but they don't give us any money to have bigger IMCs [media centers] where we could all get in, or bigger computer labs, or money to photocopy these things.

Gathering curriculum materials, conducting research, and organizing the materials and technology students needed to be successful all contributed to time spent by teachers on *Profile* implementation.

Teachers indicated that adopting a performance assessment approach also represented a major increase in the amount of time devoted to evaluation and grading. In having to spend more time in monitoring and evaluating individual student work in class, teachers had less opportunity for large group discussion or lecture. Grading time in the evenings and on the weekends also increased, as they worked to consistently grade large numbers of performance assessments. One social studies teacher reported, "Each ["Create a Nation"] project takes between 45 and 90 minutes to fully evaluate." In devising performance assessment tasks, they needed to develop self-evaluation checklists for students to complete, as well as provide their own evaluation on the same checklists. As one teacher noted, "I can't write out a checklist for every learning task for every student and still maintain the quality of instruction." Many teachers used the phrase "too much time" in discussing the time they devoted to evaluation and grading, noting in particular that they felt it decreased the time they could spend with students.

The teachers were most critical of what they perceived to be excessive record-keeping associated with performance-based evaluation and CFL reporting requirements. One teacher described this as involving completing "other checklists that have to be filled out and there are additional numbers that have to be recorded...the numbers, it's the incredible numbers, it's a hassle in terms of recording the numbers." Another explained:

> Either way it equals extra hours of work at the end of the school year when we are swamped with work anyway. This extra time the teachers need to put in does not improve the students' education in any way. I see it as busywork, paper work, unnecessary bureaucratic requirements.

In order to keep track of students' completion of various standards necessary for graduation, teachers were required to monitor whether each student was completing each standard and registering for courses that assured completion of all standards. One teacher described the process:

> We now take what was formerly a parent teacher conference day and have turned it into a registration day for next year's classes. We hope that with the parent, teacher, and student present, we won't accidentally let a kid go through grade 12 and find out that he/she is missing a graduation package that will prevent graduation. We teachers have to learn about registration and prepare materials for the conference as well as call parents and make appointments for their conferences.

Not only does record-keeping take more time, but teachers were asked to take on additional duties for record-keeping that they had not previously been responsible for.

The increased teacher time required of the *Profile* seems directly related to the finding that 53% of the teachers reported a loss in enjoyment of teaching. Teachers who had reported this were asked to explain why during the interview process. For most teachers, a primary reason for decreased enjoyment of teaching was the dramatic increase in time spent on administering and assessing the *Profile*, efforts they did not necessarily feel benefited students or themselves.

Teachers who reported a "loss of enjoyment in teaching" on the questionnaire also often explained that the loss of favorite content or projects during *Profile* implementation was partly to blame. One teacher said, "I saw myself cutting activities that students enjoy to be replaced by CFL activities that neither I nor my students enjoy." Another commented:

> I've had units that I really love teaching and really enjoy and I've had to throw those out because they didn't meet the grad standard in my class. As far as I can see the grad standard drives curriculum...The important thing in the course is to cover the grad standard and the other stuff is secondary.

In these cases, the loss of curriculum or change in curricular focus was perceived negatively by teachers.

Many teachers cited specific examples of lost content; entire chapters or units that had been cut in order to have enough class time for students to complete performance packages. Time spent in class on performance packages varied in the interview data from one to six weeks, with content being cut in order to complete the packages in all cases. One social studies teacher noted:

> We have had to cut out units on the executive branch and judicial branch so that we could fit in the packages. The executive and judicial branch are what these kids should be learning, how to make a difference in their communities through the three branches of government rather than a weak attempt to try to change something that they feel content with in the first place.

In cases such as this, teachers were not only dismayed over the loss of content but also concerned about the usefulness of the time spent instead on performance packages.

The impact of teacher preparation and resources. Two factors appear to have a strong influence on teachers' perceptions of the *Profile:* The perceived quality of their preparation for working with the *Profile*, and the perceived quality of the resources (human and material) available to assist them. (Note 3) Table 5 shows how teachers rated their preparation and resources. Tables 6 through 9 suggest a strong pattern: When teachers describe their preparation as "excellent" or "good," or when they report that the resources available to them were "excellent" or "good", they are more likely to see the *Profile* as having a positive impact on student learning and teacher instruction. However, the percentage of teachers reporting high quality ("excellent" or "good") preparation and resources is relatively low, 30% and 25% respectively.

Table 5
Teachers' Perceptions of their Preparation and Resources for Working with the *Profile* (N = 658)

Rating	Preparation for Working with the *Profile* (%)	Resources (Human & Material) for Working with *Profile* (%)
Excellent	8%	5%
Good	22	20
Fair	35	37
Poor	32	37

Table 6
Respondents' Perceptions of the Impact of the *Profile of Learning* on Student Learning by Quality of Teacher Preparation

Teachers noting increases in...	Preparation Fair/Poor (%) (Note 4)	Preparation Excellent/Good (%) (Note 5)
Students' higher level thinking	47% (Note 6)	64% (Note 7)
Students' interest in English/social studies	19	31
Students' interaction with one another	35	44
Students' understanding of grading criteria	44	51
Teacher communication with students about work quality	45	59
Students' interaction with community outside school	29	40
Quality of students' work on assignments	31	44
Students' preparation for "life" after school	26	41

Table 7
Respondents' Perceptions of the Impact of the *Profile of Learning* on Teaching by Quality of Teacher Preparation

Teachers noting increases in...	Preparation Fair/Poor (%) (Note 8)	Preparation Excellent/Good (%) (Note 9)
Teachers' preparation time for classes	80% (Note 10)	87% (Note 11)
Use of a wider range of teaching materials	45	53
Use of nontraditional assessments	34	39
Conversations with school colleagues about social studies teaching and assessment	64	71
Interaction with colleagues outside school	37	51
Interest in subject area	14	25
Enjoyment of teaching	10	18

Coordination across grade levels	37	46
Use of computer technology	32	41
Use of different teaching approaches	35	41

Table 8
Respondents' Perceptions of the Impact of the *Profile of Learning* on Student Learning by Quality of Resources Available to Teachers

Teachers noting increases in...	Resources Fair/Poor (%) (Note 12)	Resources Excellent/Good (%) (Note 13)
Students' higher level thinking	45 (Note 14)	69 (Note 15)
Students' interest in English/social studies	19	31
Students' interaction with one another	32	52
Students' understanding of grading criteria	40	60
Teacher communication with students about work quality	43	60
Students' interaction with community outside school	26	48
Quality of students' work on assignments	28	52
Students' preparation for "life" after school	22	52

Table 9
Respondents' Perceptions of the Impact of the *Profile of Learning* on Teaching by Quality of Resources Available to Teachers

Item	Resources Fair/Poor (%) (Note 16)	Resources Excellent/Good (%) (Note 17)
Teachers' preparation time for classes	82% (Note 18)	81% (Note 19)
Use of a wider range of teaching materials	44	55
Use of nontraditional assessments	33	42
Conversations with school colleagues about social studies teaching and assessment	61	79
Interaction with colleagues outside school	34	60
Interest in subject area	12	28
Enjoyment of teaching	9	20
Coordination across grade levels	33	55
Use of computer technology	31	43
Use of different teaching approaches	32	50

In the interviews, teachers noted that the most useful training consisted of helping them think about the relationships between standards and their own instruction in terms of the degree to which the instruction addressed specific standards. One teacher recalled a specific incident in which a trainer challenged the teachers to compare standards and packages:

> You would look at a package and you would say, this is unbelievable, you cannot do this in a classroom. Like you had to give speeches to authentic audiences, a variety of speeches to authentic audiences, in the package. And she would look right in it and say, does it say that in the standard? And then you'd go back and say, no. Well then does it, you know, address the standard first and then you can adapt the package. So it was her coming in and pointing out little things like that.

Other teachers believed that the success of the training depended on teachers' openness and willingness to change, as opposed to the quality of the training. Having praised the quality of the training sessions she attended, one teacher then noted that "my education about the standards came because I sought out chances to learn more, not because anyone came out to the school to present workshops." Another teacher reported that because training sessions were not required or made mandatory, teachers often did not attend: "Most of the teachers I've worked with never attended a single standards workshop."

Teachers perceived their school districts as assuming an important role in the *Profile's* implementation, particularly in terms of providing curriculum-planning workshop days and staff development support. Consistent with the survey findings, the level of variability in the quality of district training and support was perhaps one of the most important factors shaping the level and quality of standards implementation. One teacher praised her district for the "terrific job of in-service for the *Profile*. These included teaching strategies as well as meetings dedicated to informing teachers where the district stood and the process we were going through." Districts who were initially involved as pilot demonstration sites or who consistently provided support and leadership were more likely to be perceived by teachers as being successful in standards implementation.

A teacher noted the important role of district leadership:

> We were very actively involved because our assistant superintendent was very involved. To be truthful, he pushed a great deal of this onto the staff and the district, but now we are ahead of the game. Whenever I would go to the spring social studies conventions..., I always came back aware of just how far ahead our district was.

Teachers also noted the key support roles provided by curriculum coordinators in providing training, updating staff on changes, and providing resources. One of these district coordinators was praised for the amount of time she devoted to her work:

> I'm sure she was working 80 hours a week for probably about two years. And she had such a command of facts and such a big picture and she could kind of put it all in place, all these different parts and things like that. She was really kind of the glue that held it all together.

Teachers also valued the work of local curriculum coordinators who often interpreted the *Profile* in ways that were consistent with a district's own local needs, serving as a mediating bridge between state-wide accountability and local control was also praised. One teacher stated that a curriculum coordinator was sensitive to teachers' needs:

> She said, no, all you have to do then is identify those assignments that meet the standard and you keep track of those for the rubric score. Well, that's a whole different way of looking at things...her interpretation has been much better for teachers.

Often, teachers who were positive about their local curriculum coordinators were also more positive about the process of implementing the *Profile* and its impact on their own students.

On the other hand, some teachers were critical of school districts' lack of support or rigid interpretations of CFL directives. They often perceived their district leadership as more interested in pleasing CFL than in serving their own needs in terms of providing quality training. As one teacher noted, "since its inception, our district has been trying to please the state department, but not having enough direction themselves. Often times when we were being trained, they couldn't answer the questions that were being asked." Teachers pointed to this lack of clarity and leadership as a factor that limited their ability to implement the *Profile* effectively.

Another teacher perceived the district training workshops as not helping them understand ways to implement the *Profile*:

> Most of the staff came away not really feeling adequately prepped, or not feeling that they had a grasp, a full grasp of what in the world it was that these people wanted us to do. And after a number of years I still think the same thing is true.

Curriculum coordinators were often described by teachers as ineffective: "My district has one person, paid more than myself, who doesn't communicate. I've requested the goals for my grade levels and get the run-around." Another teacher commented: "Those individuals who were to direct us were confused and that led to confusion and frustration on my part." Not surprisingly, those teachers who expressed frustration with the information they received from their districts were also less positive about the implementation of the *Profile*.

Teachers also suggested that the train-the-trainer model in which district representatives attended statewide training workshops and then returned to provide district-wide training was problematic in that these representatives were often from subject matter areas different from social studies or English. As one teacher noted: "These individuals are not social studies people...they didn't know what our curriculum was, what we can add or what we can do to keep the rigor up." The lack of subject matter specificity frequently was mentioned as a significant limitation to the training teachers received.

Many of the interviewees were also critical of the lack of resources provided by districts required for standards implementation. One teacher noted the lack of funding for purchase of necessary materials "and even the money to go to another district, in order to get a sub, or even writing time...was just not available." Another teacher noted the lack of support for computer technology associated with using the Internet for research—"you have classes of 30 students and you have like 7 or 8 computers to use to access the Internet...for each student to do his or her research it takes a little longer than 5 or 10 minutes. And so you're taking days." Resource availability, like district personnel, had a significant impact on teachers' ability to implement the *Profile* effectively.

Additional teacher observations and concerns. Throughout the interviews, a number of teacher observations and concerns not directly tied to the questionnaire items emerged. Overall, interviewees were far more likely to make negative comments about the *Profile* than positive comments. Some teachers liked the idea of adopting a constructivist-based curriculum approach in theory or they strongly supported the notion of "High Standards" for all students. However, actually implementing the constructivist agenda of the *Profile* proved difficult given what teachers perceived as inconsistent direction from CFL, lack of local support and resources, public misunderstanding of the *Profile*, conservative political attacks on the *Profile*, and resistance to change. Following are some of the consistent themes that emerged.

Perceptions of the underlying philosophy of the Profile. In our interviews, teachers noted both positive and negative aspects of the educational philosophy reflected in the *Profile*. Very few interviewees were overwhelmingly negative toward the constructivist orientation of the *Profile*; most commonly teachers recognized some positives in the idea of high standards, consistency, and constructivist teaching, but had strong reservations about the "politicizing" of the standards since implementation began, or about the process of implementation. A clear disparity between the ideals of constructivist-based high standards and the realities faced by classroom teachers during implementation of the *Profile* was evident in the data.

Equity in applying high student expectations. Teachers noted the value of creating a set of high, uniform, consistent expectations that all students across the state were expected to meet. One teacher noted the need to create challenging expectations because "many teachers [are] not pushing their kids to think and be creative."

Teachers also commented that by defining standards in a consistent, uniform manner, the *Profile* provides parents and the public with some understanding of the school's specific expectations for students. They perceived the *Profile* as serving to legitimatize constructivist teacher practices for the teachers and for their students or parents who may be resisting such instruction. As one teacher noted, "With the *Profile*, we don't have to fight student and parents with comments such as 'this isn't English class' as we did in the past." Several other teachers commented that they felt that because the performance assessments were mandated by the state, and required for graduation, they were empowered to require higher quality work from students, and students were more highly motivated in completing the tasks.

At the same time, teachers acknowledged the difficulty in achieving equity across the state given the wide disparities in resources and support across different districts. They challenged the fairness of attempting to achieve the same uniform expectations throughout the state when teachers and students in poorer districts lacked the resources of richer districts. The issue of fairness was addressed multiple times by teachers reflecting on their students' access to transportation outside of the school day, computers outside of school, and parental time and assistance. Teachers in districts with a substantial number of English Language Learners (ELL) also suggested that while high standards were a good idea, the usefulness of additional high-stakes performance assessments for their students was questionable.

Many teachers suggested that the *Profile* was difficult for lower achieving students, students who are "have nots"—those who lack Internet access and/or supportive parents at home, and students who see the standards as one more "obstacle" before graduation. In particular, the *number* of required packages was seen as a concern in these instances. One English teacher described the difficulty of completing one of the literacy analysis standards for some of her students:

> The second paper was good for the advanced students, but almost impossible for the rest of our students. The teachers had to provide lists of novels that could be used for comparisons and practically outline the paper for the students. It was clearly beyond their abilities.

Teachers also suggested that if a standard was beyond students' abilities, there was not enough time to scaffold the assignment for students who required more learning time. Teachers expressed concern about students who were not able to complete the standards, asking where the time and money would be to help these students in remedial courses. One teacher drew a connection to his experience with OBE (Outcome Based Education) in Minnesota, commenting:

> This will be the same question that ultimately doomed OBE... Great plan, great thinking behind it, all sorts of great logic behind it, but the problem, the sinker for OBE was there was no one to pick up the kids who didn't make it. If we're not going to have a plan for that we need to re-think what we're doing here. And that's going to cost money. And I seriously think that's what's going to doom the program in the end. Because we're all just fine and great on reforming education until someone says whoa what will you do about this poor kid that's struggling here? And we all know they're the ones that cost money.

Concern for the impact of the *Profile* on their "less able" students was a persistent theme in the teacher interviews. Many teachers suggested that while the packages themselves were worthwhile, using them for state accountability—

their original intent—was neither realistic nor feasible.

Top-down implementation. Some teachers expressed frustration at the perceived lack of teacher input into how the *Profile* policy and framework was formulated. One teacher argued that teachers would have readily accepted the *Profile* "if it had been generated from the general teaching populace instead of the state 'experts'." Another teacher said, "I felt insulted with the fact that the state came down as if none of us were doing this and threw these packages at us." In multiple interviews, a high degree of frustration, even anger, towards the state legislature and CFL for the top-down mandate was evident.

Some teachers saw little evidence of teacher involvement in the development of the *Profile*: "The mainstream Minnesota teacher didn't have a great deal of input into the thing." These teachers complained that the state English and social studies professional organizations were not adequately consulted: "Social studies had no input into the decision-making of the standards." They also noted that much of the curriculum developed in the form of performance packages was very similar to what they were already doing. These perceptions suggest that teachers believed that they were already teaching in a constructivist manner, and that the imposition of the *Profile* was redundant and unnecessary.

Teachers were also critical of attempts to impose an external assessment system on their teaching. One teacher recommended that "each teacher develop a yearly assessment activity for their students" that would be more consistent with their own classroom methods as opposed to adopting a statewide system.

Lack of clarity and public understanding. Teachers noted that there were too many standards, and that many of the components were unclear. Many of the social studies teachers, in particular, judged the American history standard as too comprehensive and complicated for secondary instruction.

Teachers also expressed frustration with the lack of clarity in the standards statements. Teachers often had difficulty knowing how to translate the vague language of the standards and packages into their own classroom activities. As one teacher stated: "It was hard to understand. It's cloaked in jargon that very few people follow, and I know that the public doesn't understand it. I know a lot of students didn't understand it. I know some teachers don't understand it."

As with the standards statements, teachers complained about the clarity of directions and wording in the performance packages as "tedious and beyond the comprehension level of many students." One English teacher offered an illustrative example in the directions for a package requiring students to give public speeches:

> Notice that the package wants students to deliver speeches 'for a variety of purposes, situations, and audiences.' What does that mean? How do my students deliver speeches for a variety of audiences? You may think that we may be creative and haul kids to do speeches for another grade level. Kids and teachers in other grades are already busy. How about kids doing speeches for civic organizations? Great idea. But how am I going to evaluate those? Plus, in a small town, we only have a small number of civic organizations.

Teachers also noted that some of the packages were too sophisticated or elaborate for secondary students. One teacher described some of the packages as "the equivalent of at least an upper level undergraduate course, they're so complex and time consuming."

Lack of focus on teaching content. Some teachers were critical of what they perceived to be a shift away from teaching content towards a constructivist focus. They also perceived the *Profile* as representing a diminution in focus on subject matter content, particularly in terms of literature and history. One teacher said, "In our district, they don't value the content areas. It's all about process...the content has no relevance anymore." These teachers were critical of the focus on "hands-on" learning projects, noting that "the projects that we do would never be anything that I would voluntarily choose to do." Some also noted that the increased focus on a constructivist approach entails a loss of "the basics." Given their concern with the need to focus on knowledge, they perceived that the performance assessments did not provide a valid measure of knowledge as opposed to "tests [that] show what you know."

Concern with political and business influences. In the interviews, teachers were also critical of what they perceived to be the influence of political and business forces in shaping the direction of the *Profile*. Some teachers noted that the previous Governor Arne Carlson and the current Governor Jesse Ventura were unpopular with teachers and the teacher union because they failed to support what teachers believed to be adequate levels of school funding, and that these governors were seeking "payback" in the form of imposing the *Profile* onto teachers. They also believed these governors were attempting to be perceived as promoting educational reform through their support of the *Profile*, when, in fact, they were not providing additional funding. "People saw through things, like he [Governor Carlson] was claiming that he was increasing the spending on education when in reality by shifting money around and sifting things, there wasn't an increase in money." Teachers were also critical of what they perceived to be the diversion of funding from other areas in order to support the implementation of the *Profile*. One teacher noted that "all of the money that the district uses for curriculum development...has gone to write the *Profile*." Another explained, "We're in a city here with class sizes that are too big and students' needs that are too needy, and here we are spending this money, from a teacher's perspective, on things that I was doing already." In this case, the *Profile* was seen by teachers as another example of an "unfunded mandate" by the state or federal government, and many teachers suggested that the money spent on creating and implementing the *Profile* may have had a greater impact spent on reducing class sizes, providing school materials, or increasing staffing of school personnel.

Some teachers also believed that state legislators were attempting to dictate education policy without an understanding of curriculum and instruction. They noted that there was widespread misunderstanding of the *Profile* given the lack of media coverage and the lack of communication by CFL with the general population. As a result, legislators could characterize the *Profile* in ways that bore little relationship to teachers' own experiences. As one teacher noted, "The biggest thing I fear is that the legislators will start monkeying with something they don't understand. [This is] non-educators telling teachers what they ought to be doing." Another teacher lamented that "I just feel like these are more hoops for me to jump through to please politicians who know nothing about education." This notion of the *Profile* as "one more hoop" was expressed multiple times during interviews, and captures teachers' perception of the policy.

Teachers also resented the fact that the *Profile* has become "a political football," in which politicians are using the *Profile* to further their own agendas, a situation similar to that in the implementation of standards in Arizona (Smith, Heinecke, & Noble, 1999). One teacher believed that conservative legislators were using attacks on the *Profile* for their own political gain: "It's been so highly politicized that it's taken it out of the realm of education and into the realm of educational politics." Another teacher noted that the increasing role assumed by legislators in formulating educational policy was "professionally undermining—an undermining of our feeling of being professionals. Most of us have decided that both the state and the district are trying to hold teachers accountable, but that those of us already doing a good job are being punished, which is a morale destroyer." Many teachers perceived the intent of making teachers more "accountable" as an attempt to discredit teachers. As one teacher noted:

> The underlying message is that the public doesn't really believe teachers are doing their jobs.... In my department of 21, the vast majority of teachers work hard and do a tremendous job with often time-wretched resources. The few who don't are not going to change because of the *Profile*.

Questions about the political forces behind the construction and implementation of the *Profile* caused teachers to question its legitimacy.

Teachers were also critical of what they perceived as the lobbying influence of business groups in shaping *Profile* legislation and attempting to further regulate and discredit teachers. They noted that calls for increased "accountability" reflected an imposition of a business model or discourse onto education. As one teacher noted, "the attempt to lay a business model over an educational system reduced education to an accounting system rather than a human growth system." Another teacher believed that "the entire *Profile* initiative began when business leaders wanted to improve the quality of Minnesota graduates so that the profit motive might be more fruitfully pursued." Although teachers varied in what political forces they attributed the *Profile* to, they shared the perception that it was primarily a political, not an educational, initiative.

Some teachers expressed concern over the legislators' and governors' continual attempts to modify the *Profile*, noting that such changes undermined their sense of the potential long-term stability the *Profile*. One teacher complained that because the *Profile* has "gone through so many revisions, it has been nearly impossible to stay on top of the 'rules' while also educating 150 kids a day." Given their confusion about current policy, teachers often perceived the entire process as "too complex," resulting in their not knowing what to do. The continuous revisions led to a loss of support for the *Profile* over time. As one teacher noted: "I am tired of it frankly—all of the changes and repercussions on the classroom. I started out being optimistic and positive about the intent of the standards." Many teachers who had participated in the earliest phases of *Profile* implementation commented that the continual modifications led them to become disenchanted with the process, particularly because of the tremendous amounts of time they invested, only to have the "rules" changed mid-implementation.

Accountability? Some teachers argued that if accountability was truly what the CFL and state legislature were seeking to accomplish, the *Profile* would not be a scientific measure because it relied upon subjective scoring and students' work was greatly influenced by pre-teaching in their courses. One teacher explained, "I have noticed with my seventh grade son, that the way teachers approach the standard determines how much he learns from the assignment." Perceptions such as this one caused teachers to question using the *Profile* as a method of holding schools, teachers, or students "accountable."

Other teachers noted that requirements across the state were inconsistent because the legislature allowed local districts more autonomy in determining the number of standards that need to be addressed. One teacher was resentful of the fact that her district was complying with all of the standards, while other districts were requiring fewer standards, leading her to wonder about the future status of the *Profile*:

> It's frustrating for a lot of us, because we do hear that other districts are allowing kids to do things with only eight [standards] or four or none at all...we're all kind of wondering if this is going to be like OBE [outcome-based education] and just go away, whether or not it's going to stay; there's a huge dilemma.

Some teachers noted that once districts could choose their own implementation plan, that the *Profile* was meaningless:

> Actually what does the *Profile* mean now after the legislature said you can vote on it, and [name of school] and others schools are doing different things than we're doing here in [name of district]? You know, every school is doing something different. But every school's supposed to be doing the same things, aren't we?

Again, this type of inconsistency reinforced a perception of the continual change of the *Profile*, as well as the difficulty of using it as an accountability mechanism.

Discussion and Conclusions

If *Profile* proponents and opponents expect a survey such as this to prove the *Profile* "good" or "bad" for Minnesota classrooms, they will be disappointed. The results of the survey suggest the following: Some teachers perceive positive changes in teaching and learning as a result of the *Profile of Learning*. In most instances, the number of teachers reporting positive changes hovers around one-third. For example, similar to other studies, teachers report more interaction with their colleagues and greater coordination of content across grade levels (Wilson & Floden, 2001). This favorable finding is somewhat attenuated by the fact that a majority of teachers indicate that working with the *Profile* is decreasing their enjoyment of teaching and increasing their preparation time. Twentieth century U.S. education provides many examples of reforms teachers embraced but later abandoned in part because of the extensive time commitment required to implement the reform (e.g., The Eight-Year Study in the 1930s).

The results also suggest that the quality of preparation and resources provided to teachers is strongly associated with the way in which teachers view the impact of the *Profile*. Teachers who rate their preparation for implementing the *Profile* and their available resources as either "good" or "excellent" are much more likely to perceive positive changes in both teaching and learning in their classrooms. What we do not know is whether these teachers are predisposed to see the "glass half full," and those teachers who rated their preparation and resources as "fair" or "poor" are those who tend to see the "glass half empty," or whether the first group was actually involved in more substantive preparation and has access to better resources in terms of both quality and quantity. Studies of standards reform efforts throughout the country would lend support for the latter interpretation. Professional development is often the weakest aspect of implementing standards-based assessments (Herman, 1997; Kannapel, Aagaard, Coe, & Reeves, 2001; McDonnell & Choisser, 1997). A study of a statewide reform initiative in Michigan found that individual district differences, including size, structure, leadership, and readiness for change all impacted the success of standards-based professional development (Dutro, Fisk, & Koch, 2002). In the present study, less than 10% of the teachers perceived their preparation and resources to be "excellent," and over one-third of the teachers rated their preparation and resources as "poor."

A recent national survey indicated that less than half of teachers responding thought that they have ample access to curriculum guides, teaching materials, and training related to implementing their state standards (*Education Week*, 2001). Because state departments of education are often reluctant to dictate control of curriculums at the local district or school level, they may not be providing adequate guidance for strategies for implementing standards (Scherer, 2001). Conventional in-service or workshop training often provide techniques, but may not challenge basic assumptions or pre-existing beliefs about teaching (Fairman & Firestone, 2001). Major change also requires extensive resources often lacking in districts or states faced with budget cuts. Faced with the demands of everyday instruction, without time for training or curriculum-development, teachers do not acquire strategies for implementing change.

Fairman and Firestone (2001) noted a basic tension between *will*—the desire or motivation to make curriculum changes, and *capacity*—the feasibility to make such changes given time, energy, expertise, and resources. Teachers in this study indicated that when they were given extensive periods of time—often a matter of years—as well as support and resources, they were more likely to have the will to change. The fact that school districts in Minnesota at the end of the 1990s experienced marked declines in levels of state funding may mean that there is less capacity to support further *Profile* implementation.

The teachers in this study also expressed widely diverse opinions about the *Profile*. This reflects the inevitable difficulty of ever achieving consensus between educators, politicians, and parents regarding the desirability of achieving a certain set of standards (Cusick & Borman, 2002; Placier, Walker & Foster, 2002; Shannon, 2001). Such consensus presumes that all parties were privy to or were consulted on the formulation of a standards document and that these groups achieved consensus on a standards document. In a review of the implementation of a state assessment in Arizona, Smith, Heinecke, and Noble (1999) argue that instead of consensus, "assessment policy is more like a moving target that is variously constructed by political and policy actors as well as the educational practitioners who must respond to it" (p. 2). During the implementation process, different actors with different intentions enter the process with different, competing agendas (Conway-Gerhardt, 2001). Analysis of the development of the National Council of Teachers of English/International Reading Association language arts/reading national standards indicated a high level of disagreement regarding the focus, curriculum philosophy, valued instructional approaches, and strategies for implementation (Mayher, 1999; Shannon, 2001), disagreement that reflected the inevitable differences across different disciplinary and philosophical perspectives associated with teaching of language arts and reading. Given the diversity of their own beliefs and attitudes about teaching, as well as the variety of their own local teaching conditions, the teachers in this study were uneasy about any presumed consensus related to mandated state-wide curriculum and instruction.

The study does not, of course, tell us whether the *Profile* has actually prompted positive developments in classrooms; the study indicates that some teachers *perceive* the *Profile* to be having a positive impact on teaching and learning in their classrooms. A future study should be based on actual observations of teaching and learning in classrooms, as well as interviews with both teachers and students.

What are the future prospects for the *Profile?* In an insightful article on assessment-driven reform published in *Phi Delta Kappan* in 1999, Al Ramirez observed that "in state after state and school district after school district, the promise of rich assessment practices has evaporated to be replaced by the more practical and familiar approaches to testing" (p. 205). In many cases, public school officials have determined that performance-based assessments are not appropriate measures for high-stakes decisions. While the traditional testing format does not lend itself well to assessing complex thinking processes, it usually achieves high reliability and validity. Nontraditional formats such as those found in the *Profile* might be more authentic, and might give students more opportunities to demonstrate higher level thinking skills, but measurement specialists have expressed legitimate concerns about their reliability and validity. The problem becomes particularly serious when nontraditional formats are used for high stakes testing.

The conundrum, of course, is that unless the nontraditional assessment (in this case, the *Profile*) is for high stakes, school districts are unlikely to devote a lot of attention to it (Clarke & Stephens, 1996; Kannapel et al., 2000). Teachers and students are more likely to spend their time preparing for the Basic Skills Test—the "test that counts."

Aside from the technical aspects of performance assessment, it should be stressed that the *Profile* represents a significant departure from traditional views of teaching, learning and assessment—what Tyack and Cuban (1995) term the "grammar of schooling." Initially, school subjects became "learning areas," assignments and tests became "performance assessment packages," and grades of A, B, C, and D became scores of 4, 3, 2, and 1. It is not surprising that the *Profile* became quite controversial. The *Profile* illustrates Tyack and Cuban's theory that significant deviations from a community's traditional views of schooling are likely to encounter strong resistance.

As previously noted, CFL has recently adopted language that is more consistent with the "grammar of schooling." The learning area "People and Cultures" has been renamed "Social Studies," and "Resource Management" is now "Economics and Business." And although still referred to by the media and the community as the "*Profile of Learning*," CFL has renamed it "Minnesota's High Standards." The notion of "high standards," of course, is hardly controversial.

Perhaps most significant, however, is the transfer of control from CFL to the local school districts. Local control has long been a dominant theme in Minnesota's political culture. The high degree of control local school districts now have over the way in which the standards are implemented will probably defuse much of the vehement opposition to the *Profile*. On the other hand, because there is little accountability built into the system, the "high standard of performance across Minnesota" CFL had hoped to achieve is more elusive. Some districts will undoubtedly design high quality performance tasks to assess students' ability to "meet the standard," but others will address the standards at a very superficial level. The teachers who stood on the sidelines as the *Profile* was developed throughout the 1990s and predicted that "this too shall pass" may yet be proven at least partially correct.

Educational historians David Tyack and Larry Cuban offer the following observation about educational reform:

> We suggest that actual changes in schools [are] more gradual and piecemeal than the usual either-or rhetoric of innovation might indicate. Almost any blueprint for basic reform will be altered during implementation, so powerful is the hold of the public's cultural construction of what constitutes a 'real school' and so common is the teachers' habit of hybridizing reforms to fit local circumstances and public expectations. (1995, p.109)

The *Profile of Learning* as it was originally conceived deviated too much from our notion of a "real school" to become embedded in the Minnesota public school system. What remains to be seen is how teachers will shape the revised *High Standards* in their classrooms, and whether their efforts will have a substantial impact on student learning. The results of this study suggest that some variation of the *Profile* has the potential to have a positive impact on teaching and learning.

Acknowledgment

This study was funded by a small grant from the College of Education and Human Development, University of Minnesota. The authors wish to express their gratitude to the teachers who participated in the study.

Notes

1. We use the term "authenticity" as it is defined by Fred M. Newmann and his associates at the University of Wisconsin: "Authenticity is the extent to which a lesson, assessment task, or sample of student performance represents construction of knowledge through the use of disciplined inquiry that has some value or meaning beyond success in the school" (Newmann & Associates, 1996, p. 164).
2. According to Louis, Kruse, and Marks (1996), "five elements appear critical to school professional community: shared norms and values, focus on student learning, reflective dialogue, deprivatization of practice, and collaboration" (p. 181).
3. Demographic variables, such as gender, teaching experience, most advanced degree and school setting, were not associated with teachers' perception of the *Profile*. There were only two statistically significant differences between the English and social studies teachers' responses. English teachers were more likely to report that the use of the *Profile* had increased students' understanding of grading criteria, and slightly less likely to indicate that the *Profile* had decreased their enjoyment of teaching.
4. N = 441
5. N = 198

6. Forty-seven percent of the teachers who rated their preparation for working with the *Profile* either "fair" or "poor" perceived an increase in students' higher level thinking as a result of working with the *Profile*.
7. Sixty-four percent of the teachers who rated their preparation for working with the *Profile* either "good" or "excellent" perceived an increase in students' higher level thinking as a result of working with the *Profile*.
8. N = 441
9. N = 198
10. Eighty percent of the teachers who rated their preparation for working with the *Profile* either "fair" or "poor" perceived an increase in their teacher preparation time as a result of working with the *Profile*.
11. Eighty-seven percent of the teachers who rated their preparation for working with the *Profile* either "good" or "excellent" perceived an increase in their teacher preparation time as a result of working with the *Profile*.
12. N = 489
13. N = 161
14. Forty-five percent of the teachers who rated the quality of resources available to them as either "fair" or "poor" perceived increases in students higher level thinking as a result of working with the *Profile*.
15. Sixty-nine percent of the teachers who rated the quality of resources available to them as either "good" or "excellent" perceived inc
16. reases in students higher level thinking as a result of working with the *Profile*.
17. N = 489
18. N = 161
19. Eighty-two percent of the teachers who rated the quality of resources available to them as either "fair" or "poor" perceived increases in their teacher preparation time as a result of working with the *Profile*.
20. Eighty-one percent of the teachers who rated the quality of resources available to them as either "good" or "excellent" perceived increases in their teacher preparation time as a result of working with the *Profile*.

References

America 2000: An Education Strategy. (1991). Washington, D.C.: Government Printing Office.

Bakst, B. (2002, April 18). House votes to scrap profile of learning. *Minneapolis Star Tribune.* Available: www.startribune.com/stories/468/2241665.html

Bateson, D. (1994). Psychometric and philosophic problems in "authentic" assessment: Performance tasks and portfolios. *The Alberta Journal of EducationalResearch, 11* (2), 233-245.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education 5* (1), pp. 7- 74.

Clarke, D., & M. Stephens. (1996). The ripple effect: The instructional impact of the systemic introduction of performance assessment in mathematics. In Birenbaum, M., & Dochy, F.J.R.C. (Eds.). *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge,* (pp. 63-92). Boston: Kluwer.

Conway-Gerhardt, C. (2001). *Case study: Process of Implementation of K-12 academic standards and related assessment.* Unpublished dissertation, University of Wisconsin, Milwaukee.

Cusick, P. A. & J. Borman. (2002). Reform of and by the system: a case study of a state's effort at curricular and systemic reform. *Teachers College Record, 104*(4), 765-786.

Department of Children, Families, and Learning. (1998). *Graduation rule, profile of learning: Chapter 3501* (3501.0300 to 3501.1469). St. Paul, MN: Author.

Draper, N. (2000, February 3). Poll finds most teachers dislike Profile of Learning. *Minneapolis Star Tribune,* p. 3B.

Dutro, E., Fisk, M. C., & Koch, R. (2002). When state policies meet local district contexts: standards-based professional development as a means to individual agency and collective ownership. *Teachers College Record, 104*(4), 787-811.

Education Week. (2000). *Quality Counts 2000.* Available: http://www.edweek.org/srep orts/qc00

Education Week. (2001). *Quality counts 2001—A better balance: Standards, tests, and the tools to succeed.* Bethesda, MD: Editorial Projects in Education.

Education Week. (2002). *Quality Counts 2002.* Available: http://www.edweek.org/srep orts/qc02

Eisner, E. (1999). The uses and limits of performance assessment. *Phi Delta Kappan, 80*, 658-660.

Fairman, J. C., & Firestone, W. A. (2001). The district role in state assessment policy: An exploratory study. In S.H. Fuhrman, ed., *From the Capital to the Classroom: Standards- Based Reform in the States*(pp. 124-147). Chicago, Ill.: University of Chicago Press.

Firestone, W. A., Roseblum, S., & Bader, B. D. (1992). Recent trends in state educational reform: Assessment and prospects. *Teachers College Record, 94* (2), 254-277.

Herman, J. L. (1997, October). *Large-scale assessment in support of school reform: Lessons in the search for alternative measures.* CSE Technical Report 446. National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles.

Human Capital Research Corporation. (1998). *Minnesota graduation standards 1998 teacher survey.* Chicago: Human Capital Research Corporation.

Jackson, A. (2000). An investigation of teachers' perceptions concerning the implementation of performance-based curriculum and state assessment requirements throughout southeast Missouri school districts. Unpublished dissertation, St. Louis University.

Kannapel, P.J., Aagaard, L., Coe, P., & Reeves, C.A. (2001). The impact of standards and accountability on teaching and learning in Kentucky. In Furhman, S.H. (Ed.). *From the capitol to the classroom: Standards-based reform in the states,* (pp. 242- 262). Chicago: The University of Chicago Press.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994, Fall). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice. 13* (30), 5-16.

Lonetree, A. (2002, April 23). Profile of Learning survives Senate repeal effort. *Minneapolis Star Tribune*. Available: http://www.startribune.com/stories/462/2250295.html

Louis, K.S., Kruse, S.D., & Marks, H.M. (1996). Schoolwide professional community. In Newmann & Associates (Eds.). *Authentic achievement: Restructuring schools for intellectual quality.* (pp. 179-203). San Francisco: Jossey- Bass.

Madaus, G. F. & Kellaghan, T. (1993). The British experience with 'authentic' testing. *Phi Delta Kappan,* 458-469.

Mayher, J. (1999). Reflections on standards and standard setting: an insider/outsider perspective on the NCTE/IRA standards. *English Education, 31* (2), 106-121.

McDonnell, L. M. & Choisser, C. (1997, September). *Testing and teaching: Local implementation of new state assessments.* CSE Technical Report 442. CRESST/University of California, Santa Barbara.

Meisenheimer, B. (1996). *Critical attributes of teachers who have become practitioners of authentic assessment.* Unpublished dissertation, Loyola University of Chicago.

Minnesota Department of Children, Families & Learning. (2000). *Minnesota graduation standards implementation: How it looks from where it's happening.* St. Paul, Minnesota: Minnesota Department of Children, Families & Learning.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperatives for educational reform.* Washington, D.C.: U.S. Government Printing Office.

Newmann, F. M. & Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality.* San Francisco: Jossey-Bass.

Newmann, F.M., Secada, W.G., & Wehlage, G.G. (1995). *A guide to authentic instruction and assessment: Vision, standards and scoring.* Madison, WI: Wisconsin Center for Education Research.

Placier, M., Walker, M. & Foster, B. (2002). Writing the "show-me" standards: Teacher professionalism and political control in U.S. state curriculum policy. *Curriculum Inquiry, 32*(3), 281-310.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In Phye, G. D. (Ed.) *Handbook of classroom assessment: Learning, adjustment and achievement* (pp. 53-68). San Diego: Academic Press.

Ramirez, A. (1999, November). Assessment-driven reform: The emperor still has no clothes. *Phi Delta Kappan, 81* (3), 204-208.

Scherer, M. (2001). How and why standards can improve student achievement: A conversation with Robert J. Marzano. *Educational Leadership, 59* (1), 14-18.

Schomaker, M. & Marzano, R.J. (1999, March). Realizing the promise of standards-based education. *Educational*

Leadership, 56 (6), pp. 17-21.

Shannon, P. (2001). *The power of consensus: Agree, disagree, consent.* Paper presented at the annual meeting of the National Reading Conference, San Antonio.

Sizer, T.R. (1996). *Horace's hope: What works for the American high school.* New York: Houghton Mifflin Company.

Smith, M.L., Heinecke, W. & Noble, A.J. (1999). State assessment becomes political spectacle—Part I: Introduction to policy stories and policy studies. *Teacher College Record* [On-line]. Available: Doc. No. 10454

Tyack, D. & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform.* Cambridge, MA.: Harvard University Press.

Welsh, J. (2001, January 29). Officials pledge school test problems won't be repeated/safeguards 'will ensure this never happens again.' *Saint Paul Pioneer Press*, p.1A.

Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance.* San Francisco, CA: Jossey-Bass.

Wilson, S.M., & Floden, R.E. (2001). Hedging bets: Standards-based reform in classrooms. In Fuhrman, S.H. (Ed.), *From the Capitol to the Classroom: Standards-based Reform in the States* (pp. 193-216). Chicago: The University of Chicago Press.

Wiske, M.S. (Ed.). (1998). *Teaching for understanding: Linking research with practice.* San Francisco: Jossey-Bass Publishers.

About the Authors

Patricia G. Avery is a professor of social studies education at the University of Minnesota. Her research interests include civic education and teacher professional development.

E-mail: avery001@umn.edu

Richard Beach is professor of English education at the University of Minnesota. His research interests are in the areas of response to literature, media studies, and inquiry methods in secondary English instruction.

E-mail: rbeach@umn.edu

Jodiann R. Coler is currently a doctoral student enrolled in the Curriculum and Instruction program at the University of Minnesota. Her research interests include authentic pedagogy and the use of discussion as an instructional strategy.

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

This article has been retrieved 1488 times since February 23, 2003

Education Policy Analysis Archives
Volume 11 Number 8 February 23, 2003 ISSN 1068-2341

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the **EDUCATION POLICY ANALYSIS ARCHIVES**. Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

To Learn and to Belong: Case Studies of Emerging Ethnocentric Charter Schools in Hawai'i

Nina K Buchanan
Robert A. Fox
University of Hawai'i at Hilo

Citation: Buchanan, N. K. & Fox, R. A. (2003, February 23). To learn and to belong: Case studies of emerging ethnocentric charter schools in Hawai'i. *Education Policy Analysis Archives, 11*(8). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n8/.

Abstract

The fast growing charter school movement may be impeded if charter schools are perceived as a vehicle for stratifying, segregating, and balkanizing an already ethnically, socio-economically divided population. This article defines ethnocentric schools and describes three Native Hawai'ian charter schools. While they are very different in curricula and in emphasis on the Hawai'ian language and other features, they all have strong community support and a high degree of parental involvement and have access to funds available only for Native Hawai'ian programs. It may be easy to support the expenditure of public funds for ethnocentric charter schools in areas like Hawai'i where ethnic minorities have traditionally been underserved. The issues raised in this study may have broader implications for the evolution of American public education. The question is not what criteria to apply to distinguish schools of "good" choice from schools of "bad" choice. In final analysis we must ask, are schools of choice truly schools of choice, or not?

Charter schools are the most rapidly growing force within the school choice movement. Based in a quasi-market ideology that couples parental choice with school autonomy (Whitty, 1997), charter schools have strong political support from both the conservatives and liberals (Kolbert, 2000; Rees, 2000). Some support for charter schools is a thinly disguised attempt to privatize K-12 education. Others support them as a natural extension of the larger school reform movement that seeks to improve public schools for all students (Peterson, 1998). Yet others favor them as one way to avoid vouchers. One social issue that has the potential to impede the progress of charter schools is the possibility of re-stratifying, re-segregating and further balkanizing an already ethnically, socio-economically divided population (Bolick, 1997; Cobb & Glass, 1999; Crockett, 1999; Education Commission of the States, 1999; Shokraii, 1996). In this paper, we define ethnocentric schools and discuss the difficulty in arriving at such a definition, discuss historical factors that have contributed to the creation of ethnocentric charter schools in Hawai'i, describe three ethnocentric Native Hawai'ian charter schools, and suggest implications that these cases might have for the charter school movement in general.

Initially, this article was intended as a detailed study of ethnocentric charter schools in Hawai'i in an attempt to isolate common characteristics of such schools. As the reader will see, however, the extent to which such schools are subjectively self-defined led us to focus more on policy issues on which to determine the extent to which further (or expanded) support for such schools might be based.

Ethnocentric Schools

In the past, public schools focused on building democracy and assimilating ethnic minorities into a homogenized, uniquely American culture (Hlebowitsh & Tellez, 1997; Tyack, 1974). Today however, American society has become an increasingly diverse 'salad bowl' where each group remains distinct and yet contributes to a pluralistic American culture (Ravitch, 1990). The charter school movement has become one channel whereby an increasingly diverse public school population can translate demography into curriculum.

Ethnocentrism has been defined as "the feeling that one's group has a mode of living, values, and patterns of *adaptation that are superior to those of other groups"* (Columbia Electronic Encyclopedia, 2000). Ethnocentric schools have roots in the Black Power movement of the 1960's and received impetus as the Council of Black Institutions established several afrocentric schools in the 1970s to teach "children from the standpoint of their centeredness rather than their marginality" (Asante, 2002, np). Later magnet schools became a vehicle for Native American and African American educators to deliver ethnocentric education (Coffey, 2002).

Not all ethnocentric schools are the same. In general, they emphasize change in one or more of these areas: social environment, content, pedagogy, and/or language. Ethnocentric schools may provide a social environment that embraces cultural traditions and interpersonal relationship styles designed to improve student self-esteem and promote cultural identity. For example, students from the Columbus Afrocentric School strive to adhere to principles of "unity (*Umoja*), self-determination (*Kujichagulia*), collective works (*Ujima*), cooperative economics (*Ujamaa*), purpose (*Nia*), creativity (*Kuumba*), and faith (*Imani*)" (Coffey, 2002, p.3). In Hawai'i, students from Makai Charter School "*kuai I ka nu'u*" ("strive to reach the highest"). Ethnocentric schools may also change the content emphasis to reflect the contributions of their ethnic group. For Afrocentric schools that has meant teaching African history and relying on texts written about, and/or by, Black writers. For Native Americans, it has meant viewing history from an indigenous people's perspective. For Hawai'ian schools, it means "to apply the wisdom of our past to critically understand the present and create our legacy for the future" (Makai Charter School Detailed Implementation Plan). Ethnocentric schools may also adopt different pedagogies and teaching styles that they believe better match cultural teaching and learning. For example, the Native American schools may adopt a collectivistic, rather than individualistic, pedagogy that features collaboration and cooperative learning (Capozza, 1999). In addition to these changes, ethnocentric schools may incorporate native languages. Some start from English instruction and incorporate native words. Others may immerse students in their native language and assume that these students will practice English outside of school. We refer here to schools in which all or a major part of instruction is conducted in a language other than English as immersion schools.

In this study, we initially defined ethnocentric charter schools operationally as schools whose mission is the promotion and study of one ethnic group as a means of providing students with a link to their cultural heritage, sometimes including language. (As the reader will see, we eventually conclude that this definition, itself, deserves further scrutiny.) As a result of participation in such schools, students may feel increased pride and confidence in their membership in the group. Ethnocentric schools employ teaching strategies that are congruent with the learning styles and preferred ways of processing and acting on information that reflect the cultural heritage of their target population. The stated goal of such schools is to use these as vehicles for generating improved performance from students underserved by traditional schools. The definition of Native Hawai'ian varies depending upon the organization proffering the definition. For example, as a criterion for service eligibility, both the Kamehameha Schools (a multi-billion dollar private academy funded by the estate of Bernice Pauahi Bishop for the "education of the children of Hawai'i") and the Office of Hawai'ian Affairs (established by the State of Hawai'i to manage funds held in trust for Native Hawai'ians) define as Native Hawai'ian any person who can prove Hawai'ian ancestry, while the Department of Hawai'ian Homes (another agency which assists citizens of Hawai'ian ancestry to take up residency on lands that were originally held by the Hawai'ian monarchy) requires that a person have 50% blood quantum to be considered Native Hawai'ians. Throughout this paper, we use Native Hawai'ian and part Hawai'ian to encompass any person of Hawai'ian ancestry.

Ethnocentric Schools and the National Charter School Movement

National and state charter school reports provide data about the ethnic/racial and 'at-risk' distribution of students in state or chartering districts (Center of Urban and Public Affairs at the University of Texas - Arlington, Texas Center for Educational Research and Center for the Study of Educational Reform at the University of North Texas, & Center for Public Policy at the University of Houston, 2000; Nelson, et al, 2000; Public Sector Consultants Inc., 2001; Wells et al., 1998). So, for example, a Michigan study reports "the percentage of minorities in the study-area charter schools is higher than in both the state as a whole and the traditional public school districts in which the charter schools are located" (Public Sector Consultants Inc., 2001, np). However, Crocket (1999), in her study of California charter schools, found that charter schools were 63% Whiter than their sponsoring districts. Other researchers note that aggregate data reported in such national and state studies may actually mask ethnic stratification (Berv, 1998; Cobb, Glass & Crockett, 2000; Fusarelli, 2000). These reports focus on the issue of White flight and skimming the brightest students into elite schools. They, however, fail to explore the impetus for, and dynamics demonstrated in, the purposeful creation of ethnocentric schools of choice for indigenous students, students of color and minority populations.

Native Hawai'ian Charter Schools

Hawai'i is the only single -district state in the United States. An elected Board of Education (BOE) appoints the

Superintendent of Schools, serves as a policy-making governing body and establishes priorities for the allocation of state funds subject to the political realities within which it must operate. In 1994, pressure from school reform advocates resulted in legislation that allowed for 25 existing schools to convert to student-centered schools, specifically avoiding the term charter school. Student-centered schools were given some budgetary control and the opportunity to request waivers of some rules and regulations from the Department of Education (DOE). Only two schools out of 253 chose to convert.

Legislation enabling twenty-five New Century Public Charter Schools including start-ups, school-within-school programs and whole school conversions was passed in April of 1999. By September of 1999, over thirty groups had submitted letters of intent to become charter schools. The new law clearly attracted two distinct populations whose needs were not met by the current system. The first group consisted of Native Hawai'ian communities (50% of the letters of intent from throughout the state). The second overlapping group consisted of programs and groups from the neighbor islands (60%). In Hawai'i the central administration of the DOE and most other government agencies are located on the island of Oahu, geographically small but with the largest population. The other inhabited islands are often referred to as the neighbor islands. The primary reasons for starting charter schools in Hawai'i appear to be autonomy from a distant center of control and the desire to serve a neglected special population of Native Hawai'ian and part Hawai'ian children.

Although some form of state governing board for education exists in all fifty states, the District of Columbia and Puerto Rico, the state of Hawai'i is the only single statewide school district lead by a single state superintendent responsible for all public k-12 education accountable to a single Board of Education. This reflects the history of the state that, until as recently as 1955, was controlled by an elite, primarily White oligarchy of plantation owners. The vast majority of citizens are descended either from displaced Native Hawai'ians or from populations imported from Japan, China, Korea, Portugal, and other countries to work in the fields (Langlas, 1998).

Both the organization and the philosophy of the Hawai'i State Department of Education has tended to reflect its heritage, with highly centralized decision-making, dependence upon rules, regulations and rubrics, and the pervasive view that the central administration knows best (Dotts & Sikkema, 1994). The Department of Education oversees two hundred fifty-three schools, one hundred eighty thousand students and sixteen thousand employees (Office of the Superintendent/Planning, Budget, and Resource Development Office, 2001). This is complicated by recent dramatic increases in the number of at-risk students. Since the 1990-1991 school year the total school enrollment has grown by 8.3% while the number of students who receive free or reduced lunches has grown by 66%; are identified as in need of special education services by 97%; and have limited English proficiency by over 70%. Only 49.4% of the school population is considered not disadvantaged (Office of the Superintendent/Planning, Budget, and Resource Development Office, 2001). This comes at a time when Hawai'i leads the nation in unemployment and 31% of Hawai'i's children live in families where no parent has full-time, year round employment (PRB/KIDS COUNT, 2002). Hawai'i has also been cited as having one of the largest average school sizes in the nation and the lowest annual increase in spending for education of comparable states (Office of the Superintendent/Planning, Budget, and Resource Development Office, 2001). It is not surprising that this has resulted in a school system that is given a grade of C or less by 73.9% of the people in the Hawai'i Opinion Poll on Public Education 2001.

Adding to the stresses placed on Hawai'i's school system has been a growing realization on the part of indigenous Hawai'ians that society in general (and the school system in particular) was neither meeting their needs nor sensitive to their culture (Buchanan, 1998). Native Hawai'ians make up 0.8% and part Hawai'ians 17.5% of the population of Hawai'i (Schmitt, 1998). A variety of structures unique to the state of Hawai'i originated with the forcible overthrow of Hawai'ian Queen Liliuokalani. Hawai'ian as a medium of instruction in the public schools was banned in 1896. Beginning in the late 1960s a cultural renaissance began that resulted in the revival of dance, music, cultural practice and language. In 1978 the Hawai'i Constitutional Convention declared Hawai'ian to be one of the two "official" languages of the state and mandated the provision of educational programs in Hawai'ian language and culture. By 1984, a determined group of Hawai'ian speakers successfully launched the first Punana Leo and Kaiapuni Hawai'i program that created preschool language immersion programs (Kapono, 1998). Legislation in 1986 expanded the immersion program k –12 with the result that immersion programs became, for the first time, a responsibility for the already-overburdened public education system.

Like many social movements, the demand for educational reform initiated in 1983 by "A Nation at Risk,", reached Hawai'i (in the middle of the Pacific, 2500 miles from its nearest neighbor) considerably later than on the mainland. Indeed, it was not until 1989 that the Hawai'i Legislature directed Hawai'i's Department of Education to design a School Community Based Management approach to incorporating parents, community leaders and teachers into educational decision-making. The SCBM program, though relatively short lived in Hawai'i, was notable primarily for two features; the extreme limitations placed upon genuine efforts at reform by the central administration and the incorporation of traditional Hawai'ian values such as *lokahi* (harmony), *kokua* (helpfulness), *laulima* (cooperativeness) by local groups seeking reform. The community-based decision-making that was allowed required consensus and emphasized avoidance of embarrassment over substance (Hawai'i State Department of Education, 2002).

In 1995, the Legislature, recognizing that the Board of Education was unable or unwilling to bring about genuine school reform, passed legislation empowering local groups (under strict limitations) to form "Student Centered Schools" which, while public in most ways, were allowed limited local autonomy under a local school advisory board. These were to become the precursors of the Charter School movement in Hawai'i. Although, typical of Hawai'i, the two conversion schools which were established under this legislation were located in two of the most economically elite areas in the state. The population of students at each of these two schools identified as indigenous Hawai'ians or part Hawai'ian are only 20% and 13% respectively while the largest ethnic populations at these schools are 59% White at one and 42% Japanese at the second. Although nominally locally controlled, these schools operated with virtually the same faculty and school level administration, followed almost all Department of Education curricular, financial, and personnel procedures, remained in the buildings which they had previously occupied and, for many, were distinguishable from traditional public schools in only superficial ways.

In 1999, when the Legislature, abandoning even more aggressively its efforts to bring about change within the state educational system, passed Hawai'i's first real Charter School empowering legislation, the two Student Centered Schools became Hawai'i's first "New Century Charter Schools." Soon, with the encouragement of the Federal Charter School Program, more than 30 groups prepared to compete for the remaining twenty-three charters permitted under the law. This paper, however, focuses on an unexpected (by some) phenomenon which emerged as the various planning groups developed Detailed Implementation Plans (DIPs) in pursuit of the much-sought-after Charters and, thereby, some freedom from the central Department of Education.

Sixty percent of the founder groups were located on islands distant from the state capital (Honolulu) located on the island of Oahu. And, even more striking, 50% identified themselves as being ethnocentrically Hawai'ian. For some, this meant a focus on the language and, indeed, five charter schools are currently conducted all or in part in Hawai'ian (referred to in Hawai'i as "immersion" schools). For others, the focus was on Hawai'ian culture as a nurturing environment (absent, the argument went, in traditional public schools) within which disadvantaged students of Hawai'ian ancestry were more likely to learn. Still a third group sought to apply Hawai'ian epistemology as a means of conveying both traditional and Hawai'ian subject matter.

What began as a law to empower the creation of a limited number of charter schools became a strong force for ethnocentric education in the state. This paper examines three Hawai'ian ethnocentric charter schools on the island of Hawai'i (referred universally as the "Big Island" to distinguish its name from that of the state).

Methods

We selected three self-defined ethnocentric charter schools operating on the Big Island. Although all three share certain common characteristics, their significant differences permitted examination of a variety of different approaches. The first—a total immersion Hawai'ian language charter school—is situated on the campus of an existing traditional Department of Education school that conducts its classes in English. While the charter school classes are conducted in the Hawai'ian language, its curriculum and structure reflect traditional knowledge and skills . Having previously operated as a school-within-a-school on its campus, it might appropriately be considered a conversion charter school in many ways. This school—called Makai Charter School (MCS) for this report—is located in a community with an extremely high percentage of Native Hawai'ian residents.

The second, located some sixty miles from MCS, came into existence as Koa Public Charter (KCS) school by combining three components: a 9th – 12th grade school-within-a-school sited on a local traditional high school campus , a pre-existing private primary (preschool – 4th grade) total immersion Hawai'ian language school and a newly-created 5th through 8th grade middle school. Classes are conducted in English, although the Hawai'ian language is heard frequently from both students and teachers. At the time of its formation as a public charter school, the school moved out of its previous site on a high school campus and might appropriately be seen as a start up charter school. The ethos of the school reflects the founders' belief that traditional Western education has both failed Native Hawai'ian children and has eroded traditional Hawai'ian value systems.

The Hilo Charter School (HCS), a start up charter school underwritten by an existing Foundation dedicated to the preservation of Hawai'ian culture and values, lies somewhere between the other two in its educational philosophy. It is conducted in English (although, as above, the Hawai'ian language may be heard everywhere throughout the school) and, while heavily devoted to the "Hawai'ian way of life," is less negative about the perceived failure of traditional Western education. Of the three schools, HCS draws most heavily on its connection to the local community and to the Hawai'ian elders (kupuna) associated with its sponsoring Foundation. It is located on a fourteen-acre site provided by the Foundation and looks forward to significant construction of classrooms and instructional facilities.

In Hawai'i, as a condition of being granted a Charter by the state Board of Education, each school must submit a Detailed Implementation Plan (DIP) setting forth the philosophy, pedagogy and organization of the proposed charter school. We, co-founders of the University of Hawai'i Charter School Resource Center, have followed the development of charter schools in Hawai'i from the start. We began by studying the DIP from each of the three schools, with particular focus on statements about ethnic identity and the values of ethnicity and the use of the Hawai'ian language. From this, we identified the following questions to be directed to the schools:

- What historical factors contributed to the creation of this ethnocentric public charter school?
- What major changes stimulated or discouraged the creation of your school?
- In what way does the actual operation of the school reflect the ethnocentric goals of your mission?
- How are resources (including physical space and human resources) funded? How accurate were the initial estimates of school costs? How are decisions made when funds are insufficient to cover all costs?
- What is the interaction between physical space and school mission?
- What implications do you think your school has for the charter school movement in general?

In addition to interacting closely with these schools in the pre-start-up period and reviewing the DIPs, we visited each school at least once to conduct formal interviews. Students were observed at work and at play. Members of the staff were interviewed. Since two schools have been operating for less than a year and one for two years, objective evaluation of educational effectiveness is not yet available. This report seeks to address (1) the extent to which the affect of the entire school reflects its ethnocentric nature and the goals of its founders, (2) the degree to which the founders and members of each school community have been able to create an institution which reflects the aspirations in their DIP and (3) anecdotal evidence of the extent to which the Department of Education has supported or impeded school development.

The results reported here are less designed to be exhaustive than to identify public policy issues related to ethnocentric charter schools for which further study is indicated.

Case 1 – Makai Charter School

Finding the office of Makai Charter School (MCS) is a challenge. No signs distinguish it from the other classrooms and offices that house both a k-6 regular DOE school and Makai charter school. According to the most recent School Status and Improvement Report (2001), 24.2% of the students at the DOE school are Hawai'ian and another 66.3% part Hawai'ian for a total of 90.5%. Hawai'ian/part Hawai'ian student enrollment at Makai CS is above 94%. This can be compared to the two nearest DOE elementary schools whose student bodies are 46% and 33% Hawai'ian/part Hawai'ian respectively.

The old wooden structures appear to need refurbishing, and the hallway that leads to the MCS office passes a storage area of broken desks and other miscellaneous furniture. The office is a semi-underground area with painted pipes and exposed wiring running along the ceiling. Despite the less than ideal physical surrounding, the principal, secretary, and clerk are productively engaged at their computers and phones preparing for the 8:15am to 2:15 pm school day to begin. They interrupt their normal routine and, joined by a young counselor, all enthusiastically greet us and answer questions with pride about their school.

Originally, MCS was a Hawai'ian Language Immersion school-within-a-school (SWIS) established as continuation of a Punana Leo language immersion preschool, part of the Native Hawai'ian cultural renaissance. As a school within a school, there was tension between the regular DOE and SWIS staff. Becoming a charter school meant new autonomy and self-determination. It empowered the staff to make more decisions about how and what to teach as well as how to schedule their time. This is the first year of operation as a charter school. The 149 students are grouped into seven classrooms (grades k/1, 1/2, 3/4, 4, 5, 5/6, 6). Even though efficiency concerns have forced the school into multi-grade groups , one class still houses 31 students. MCS has adopted a trimester calendar and extended school day that facilitate language learning by replacing summer vacation with fall, winter and summer inter-sessions with the longest a one month summer inter-session. This is also designed to counteract the effects of Hawai'i's short school year, the shortest of any state.

The MCS classrooms contain typical k-6 colorful posters and student work evident on the walls, hanging from the ceiling and stored on shelves around the room. Before these students enter the school for the day, they gather outside on the lawn and ask permission to enter the school. The principal's chant gives them permission to enter and reminds them of their responsibilities to learn and behave. This Hawai'ian protocol is followed each day. In one class a 6th grader stands at the front and spells one of the weekly words and then reads his sentence to the class. In another, the teacher reads a story, and yet in another students work independently on math worksheets. All the teachers at MCS are licensed by the Hawai'i State Teacher Standards Board.

The MCS vision is "Inspired by our past. Empowered by our identity. Prepared for the future." Its mission is to be a "culturally-based indigenous k-6 Hawai'ian Language Immersion school…" that "promotes Hawai'ian ways of knowing to strengthen and revitalize a Hawai'ian identity…" in "experiential-based Hawai'ian learning environments" (DIP). The school's goals for students are the development of literacy and communication skills, personal and social responsibility and thinking and reasoning skills. The Experiential-Based Activity Model (EBAM) designed to help students explore interdisciplinary problems and practical applications of knowledge and information (Moersch, 1994) is one of the main strategies employed at MCS. The Hawai'i Content and Performance Standards II (HCPSII) that are mandated for use by all DOE schools guides the curriculum at MCS along with a commitment to Hawai'ian language immersion, culture and values. Since the Hawai'i Assessment Program has no tests translated into the Hawai'ian language, the school is considering whether to begin formal English instruction earlier (currently being in 4th grade) and thus become more fully bilingual. There is a tension between helping Native Hawai'ian students be successful in the modern world and restoring the native language that may not contribute to economic or social growth.

To assess student progress toward meeting HCPSII, the school has adopted Work Sampling System's developmental checklists (Rebus Planning Associates, 1994) to replace traditional repost cards. In addition, MCS has adopted the Hawai'i State Superintendent Accountability Design and the National Study of School Evaluation as part of a school accountability system in addition to adopting a sound fiscal responsibility plan.

Students of any ethnicity may apply to MCS but the full immersion curriculum clearly places practical limitations on entering students. Students who enter without Hawai'ian language preschool experience, have experienced only English instruction in grades k –5, or do not speak Hawai'ian at home are less likely to succeed in the immersion program. The first item on the application asks parents to:

Please initial before the box ALL that are applicable:

[ ] The student is currently attending {language immersion school within a school} and will be returning in SY 2002-2003.

[ ] The student is a sibling of a returning student.

[ ] The student is a transfer student.

[ ] The student has no Hawai'ian language background. (For Kumu info only.)

The sense of community and inter-generational continuity is an essential part of the Hawai'ian culture. Indeed, the immersion school movement has, from its very beginning within the DOE, provided for mandatory parental involvement. This continues to be one of the most striking characteristics of Hawai'ian ethnocentric charter schools. Before selection is made, parents must complete an agreement to participate form that states:

> I (We) understand and agree that my (our) child(ren) will be educated through the medium of the Hawai'ian language. I (We) understand that one (1) formal English class will be introduced in the 4th grade and will continue through 6th grade.
>
> I (We) understand and agree that I (we) actively support my (our) child's (ren's) learning through the availability of Hawai'ian language classes, self-help books with cassette tapes and pre-taped video coursework if I (we) are not yet fluent in the Hawai'ian language.
>
> ...enrollment is contingent on space availability and acceptance of the charter school's vision, mission and goal statements.

In addition parents must agree to attend at least three parent meetings a year, two student activities a year and contribute two hours per month in volunteer work for the school. Currently 6% of the students are non-Hawai'ian. Many students come from out of the geographic area.

Case 2 – Koa Charter School

Upcountry Hawai'i offers lush landscape, almost constant wind and alternating sun and clouds and rain. Koa Charter School (KCS) is off the main road unannounced by signs and situated on 6 acres of agricultural land lent to the school by the Department of Hawai'ian Homelands and another 4 acres used in collaboration with the YMCA about 10 miles away. The first site has a house that has been converted into offices: the nerve center of the school. Approaching the office one passes a large warehouse, two large and one small white-tarped quonset huts. The warehouse serves as a computer and technology lab, library, lunch distribution site and instructional space. The two quonset huts are divided into two classrooms each with bookshelf dividers between the rooms. Usually two adults work with eight to twelve students on a variety of skills each morning. The huts have cement floors and slanting sides that have wire strung to hold brightly colored student work. Whiteboards are set along the walls and bright Hawai'ian cloth is suspended from poles that support the hut. One side is usually open but can be closed by fastening tarps at both sides. Most students are in multi-aged groups. School always starts with Hawai'ian protocol like the one described above at MCS.

One formal English class we visited consisted of eight 2nd – 6th graders who were all on about the same level studying English sight words, copying them on one page and using them in sentences on another page. Some students have come to KCS from language immersion programs and others from regular DOE schools, so skills in English and Hawai'ian make instruction a challenge for KCS teachers. Children from the two classes in the hut gathered outside on the lawn before lunch to *pule* (pray); an important activity at any gathering of Native Hawai'ians.

The KCS vision is to become a comprehensive education and service center for Hawai'ians of all ages. The school evolved from two antecedents: one a 9th-12th interdisciplinary academy school within a school, and the other a Hawai'ian Immersion preschool – 4th grade. These combined and included the intermediate grades to form a k-12 bilingual charter school. Its founders believe that indigenous peoples have the right to design and control their own education and further that Hawai'ian people can be successful in the 21st century without giving up their culture, language and traditions. The founders believe that Hawai'ian culture has deteriorated because of Western philosophy, religion and laws that advocate that man subdue the earth for profit and personal gain rather than exist as stewards of the land. Another impetus for the school was a desire to slow the out migration of Hawai'ians and develop an economy that would allow Hawai'ian graduates to remain in the islands. The KCS vision is "strive to reach your highest potential" (DIP). Students and staff at KCS are expected: to love one another, take care of their responsibilities, give and receive help, and be thankful for what they have.

Eighty-eight-percent of the school's 150 students are Native Hawai'ians/part Hawai'ian. With Federal funding in addition to the per-pupil DOE allotment, the school has been able to operate with 51 'teachers', some licensed and others educational aides or specialists. In addition to grants specifically for Native Hawai'ian education, KCS receives Title 1 funds. They serve 15 (10% of the KCS population) special education students and provide gifted and talented activities through a federal Native Hawai'ians grant for all students. KCS boasts an attendance rate of 97%, one of the highest in the state. By comparison, the nearest DOE school, a middle school, reports a population of 34.3% Hawai'ian/Part Hawai'ian. 13.7% of the students at the DOE school participate in special education programs and average daily attendance is 91.4% (Department of Education School Status and Improvement Reports, 2000-2001).

The KCS curriculum is a balance of culturally driven and standards based strategies that emphasize: reading, writing and communication in both Hawai'ian and English; the ability to apply math and science; the ability to access, evaluate and use a variety of technologies; to apply critical thinking and problem solving; the mastery of academics, culture and workplace skill; and the development of work ethics necessary for economic self-sufficiency.

Originally, the plan was to have two multi-aged groups of students with approximately 25 elementary, 25 middle and 25 high school aged students in each group. Each group was to remain together for a full year and work on theme-based interdisciplinary projects related to Hawai'i that had social significance for Hawai'ians. Through the projects students would demonstrate essential competencies and performance standards including technology and career explorations and would contribute to sustaining healthy economy in the community. Each group would spend two days each week at a lab site; either the Hawai'ian Homelands site in the rainy, forest or the dry-land ocean site. They would spend the other two days documenting their projects. On Fridays, students would participate in Student Development Workshops where they would explore careers, engage in community service and work with mentors in the community on personal development such as health and fitness. Everyday Monday through Thursday students would: 1) use the Hawai'ian language for opening and closing protocol; 2) have 20 minutes of Total Physical Response that emphasized both the Hawai'ian language and physical fitness; 3) do a problem of the week to assure that students engage in problem-solving and record their work in a journal; and 4) engage in sustained uninterrupted reading for pleasure.

However, when the school opened, the teachers found that it was difficult to meet the needs of k-12 students within a single group. They modified the grouping so one group consist of k-5th grade and the other 6th – 12th grade students. Mornings are used to develop basic skills. Each student is pre tested in reading, writing. and math and multi-aged group according to skill. In the afternoon, these same groups engage in projects. This unanticipated change in the basic format of instruction is too new to allow either the school or the authors to evaluate the extent to which it re-defines the original goals of the founders. It does, however, provoke some suspicion about the ability of even the purest educational philosophy to withstand educational reality and the day-to-day pressures of dealing with undereducated children.

Case 3- Hilo Charter School

Seventy students, 5 core teachers and many volunteer community members conduct classes on an undeveloped fourteen-acre site and several subordinate sites (all located within about a mile of each other) in a community with one of the highest populations of Hawai'ians (and, not coincidentally, one of the lowest economic levels) in the state. One hundred percent of the students at HCS is Native Hawai'ian. Ironically, the location of the sites (some of which front directly on the Pacific Ocean) makes the land on which the school sits some of the most valuable in the state. As the beneficiary of its sponsoring Foundation, Hilo Charter School will, in the future, enjoy facilities beyond the reach of many schools. For now, however, the site is largely undeveloped and classes are held in various structures ranging from a large undifferentiated room in a brand new community hall to an open air structure constructed of pipe frames and agricultural tarpaulins. Students play in an open field combining breathtaking beauty with a total lack of recreational facilities. The campus, as is the case at KCS and several other ethnocentric Hawai'ian schools, is heavily planted with indigenous plants; most of which have economic, cultural or spiritual significance to the Hawai'ian people . Agriculture (and aquaculture), geneology, and navigation/astronomy form the core of the educational experience at KCS and in Hawai'ian culture, which places emphasis on the relationship of people to each other and to the land and the sea.

The school conducts classes for children from 7th through 12th grade. A pre-school operated separately by the Foundation occupies a site at the far end of the campus. A separate large room with few partitions and no interior walls serves as school office, staff workroom, lunchroom, meeting room, etc. Multiple activities are conducted in the single-room community hall. Four or five classes simultaneously meet in corners of the room. One portion of this large space is given over to fifteen new lap top computers that sit on low, Japanese-style tables and are in heavy use by students. The contrast between the rustic nature of the site and the enviable array of technology is striking.

The relative quiet and calm demeanor of the students at the school, even during lunch and recess, was noteworthy. With few teachers in evidence and no intrusive adult supervision during recess free play, students seem happy and self-directed. The end of recess was announced by the blowing of a conch shell (a traditional Hawai'ian call) by one of the teachers. It was interesting to observe the relative ease with which the students finished their field games (most involving a dodge ball-type game) and returned without complaint to their lessons. Familiar boy-girl posturing often observed on 7th -12th grade campuses was not in evidence. Lessons are conducted with the students sitting on the floor (in traditional local fashion, everyone removes his or her shoes at the door). In some cases, the students were arrayed in semi circles around the teacher. In others, the classes were obviously more diffuse, with the students reclining at short-legged tables while the teacher moved from group to group.

Observers used to traditional classrooms might find the room unsettling. There are no chairs and few tables. There was not a blackboard in sight. No walls separated one class from the others. The room was, however, surprisingly not chaotic. Noise level was at a minimum because there was very little off-task talking between the students in different groups. In fact, it was difficult to find a student whose face was not intently directed either toward the teacher (in those classes where teachers stood at the front) or at his or her work (in those classes where the teacher moved from group to group).

Lessons cover traditional topics (ultimately state legislation requiring evidence of adherence to Hawai'i's Performance and Content Standards both motivates instruction and limits the extent to which innovation can occur) but there was obviously a project-based flavor to the classes. One group, for instance, combined art, science and language as they worked on landscape plans for the campus. Other groups study the Hawai'ian approach to astronomy, their relationship to the land and the sea, ecology and Hawai'ian health. Apart from the physical arrangement, the classes did not look substantially different from those in most schools. The difference was in the affect; in the expectations (and proffering) of respect that Hawai'ian children traditionally give to adults.

The Hawai'i charter school law does not require it, but one of the Director's first comments was that all teachers at HCS were licensed. He gave us a tour of the campus and then sat down for an extensive interview and discussion. "Hawai'ianness" at this school manifests itself primarily in two ways: focus on Hawai'ian-related, project-based instruction and respect for the Hawai'ian environment and community with which the school closely relates. Subject matter selection is driven largely by the Hawai'i Performance and Content Standards and is, therefore, not that dissimilar from other, non-ethnocentric schools. It is not clear which is cause and which is effect; the traditional manner in which all of the teachers have been trained or the fairly traditional pedagogy.

"Regular" classes for students take place both on the main campus and on two nearby sites; one for agricultural projects and the other for ocean-related activities. Teachers teach from Monday through Thursday and meet together on Fridays. On Fridays, the school imports local resource persons to provide an enriched elective environment with heavy emphasis on Hawai'iana (hula, fishing, canoeing).

Perhaps the most significant evidence of the ethnocentricity of the school, as reported by the Director, is its situation within the local, Native Hawai'ian community. Relations with parents and community leaders are very close, with parents and (importantly in the Hawai'ian community) grandparents being seen as members of the holistic educational team. The importance of kupuna (Hawai'ian elders) is infused throughout the school.

HCS is eligible for a variety of federal and private supplementary funds without which they would not survive. HCS is also fortunate because its sponsor Foundation has access to significant land (of the three schools studied, Hilo is the only one with its own campus for which it can make permanent plans) and sources of funding. This means that HCS can focus its energies on educational development; not finding and funding facilities.

Discussion

The questions that originally motivated this study were modified to be more consistent with the Hawai'ian tradition of 'talk story'(Dotts & Sikkema, 1994). The respondents were obviously very proud of what they had created and "wanted to talk about what they wanted to talk about;" firmly but persistently resisting efforts to re-focus. Upon reflection, we were reminded that Native Hawai'ians rely on verbal, rather than written, history (Langlas, 1998). We concluded that the best course of action was to let the study take us wherever it went. The discussion which follows attests to the value of that approach.

The three schools studied were at once significantly different and strikingly similar. Each school is a Hawai'ian ethnocentric charter school largely on the basis of self-definition. All three DIPs and sets of promotional materials describe the unique and fragile nature of Hawai'ian language and culture. MCH writes of "Hawai'ian ways of knowing" and "experiential-based Hawai'ian learning environments" while KCS emphasizes "the Hawai'i indigenous people culturally-driven educational milieu" and HCS advocated "rebuilding a Hawai'ian intergenerational community." Although the literature contains descriptions of Hawai'ian epistemology (Meyer, 2001) and attempts to describe the Hawai'ian worldview, we observed few attempts made by these ethnocentric schools to define their own terms. One is left with the sense of "we know it when we see it." However, both state and federal statutes place severe limitations on the ability of a publicly funded charter school to discriminate in any fashion. Therefore, the actual extent to which the ethnicity of any Hawai'i public charter school can be identifiably Hawai'ian can be attributable to location (schools located in ethnically identifiable neighborhood tend to draw from the locality; particularly in regions with limited public transportation) and parental selection (not surprisingly, an emphasis on Hawai'ian language and culture is disproportionately of interest to ethnically Hawai'ian families). Nevertheless, the populations of the three schools are overwhelmingly composed of students who identify themselves as Hawai'ian or part Hawai'ian (MCS - 94%, KCS - 88% and HCS -100%).

Each school clearly identifies itself as a member of the Hawai'ian ethnocentric school subset of Hawai'i public charter schools. In 2000, the leaders of one of the three schools founded Na Lei Na'auao, an organization of identifiably Hawai'ian ethnocentric charter schools that has grown to include 12 such schools in the state. In addition, a statewide Hawai'i Association of Charter Schools (HACS) with representatives from ethnocentric and non-ethnocentric charter schools meets periodically to liaise with the Department of Education and to lobby for improvement (or lobby against deterioration) of state charter school enabling legislation. However, Na Lei Na'auao remains as a clearly identifiable "ethnocentric schools only" organization. A bill passed by the Hawai'i Legislature in April, 2002 would allow non-profit organizations such as Kamehameha Schools to run conversion charter schools with augmented operational funds from their non-profit organization in geographic areas that have large populations of Hawai'ian/part Hawai'ian students. Kamehameha Schools was founded at the beginning of the last century by the estate of Princess Bernice Pauahi Bishop for the "education of the children of Hawai'i." It has grown to a multi-billion dollar private educational institution serving only children of Hawai'ian ancestry. The entrance of this institution into the establishment of Hawai'ian ethnocentric charter schools may significantly change the financial and political balance of the charter school movement in Hawai'i.

While Wells, Lopez, Scott and Holme (1999) identified a composite category of California charter schools that were termed "urban, ethnocentric, and grassroots charter schools", the ethnocentric charter schools in Hawai'i share characteristics that do not fit this category. Hawai'i schools are primarily rural and focus on the indigenous Native Hawai'ian culture. The three cases reported here can be described in a number of dimensions: physical environment, personnel, sources of funding, relationship with parents and the community, curriculum structures, pedagogy and language, and educational goals.

The physical environment seemed determined by whether the charter school was a new start-up or a conversion program; true for non-ethnocentric conversion and start-up charter schools as well as the ethnocentric schools. In our sample, the conversion SWIS, MCS, remained in a traditional classroom setting which appeared to contribute to a more traditional delivery of instruction. The two start-up charter schools, KCS and HCS, were challenged by the need to create new physical spaces and adapt to non-traditional classroom spaces and these uncommon settings appeared to make it possible to try innovative programs in more natural settings. So, for example, students at HCS could spend their afternoon classes at the beach studying water quality or conducting reef fish surveys as part of their course of study.

While many of the teachers and members of the staff at the three charter schools are Hawai'ian, the percent of teachers of Hawai'ian/part Hawai'ian ancestry was lower than that of the school population. At both MCS and HCS, all of the teachers were licensed and the school leaders valued the credibility that this brought. At the more rural KCS, the leaders used federal, state and private grant funds to hire 51 "teachers." Of these only 5 are licensed. KCS is currently using federal grant money to fund an alternative Hawai'ian teacher education program that will enable them to grow their own licensed teachers.

At this time every non-ethnocentric charter school in Hawai'i has experienced broken fiscal promises and bureaucratic interference. Indeed, three start-up charter schools have litigation (Note 1) in progress against the state and the Board of Education asserting that charter schools receive substantially less money than other public schools and, more specifically, less money than they were originally promised. It is notable that none of the three are Hawai'ian ethnocentric charter schools. While it is not the purpose of this article to examine the validity of these claims, it is unarguable that the financial condition of Hawai'i's charter schools is bimodal. Ethnocentric charter schools are surviving; the rest face bankruptcy. The three schools in this sample all have outside funding from federal grants specifically earmarked for Native Hawai'ian education, health and environment and several state-based Native Hawai'ian foundations. In addition, the Hawai'ian charter schools have access to land and in some cases existing buildings that can be or are being used to house schools.

While all schools recognize the value of parental involvement in their child's education, the Hawai'ian charter schools each have a characteristically Hawai'ian commitment to parental and community connection to the school. KCS envisions "a comprehensive Native Hawai'ian learning center or *kauhale* which can address the educational and cultural needs of all stakeholders from womb to tomb." HCS is determined to be a part of the Hawai'ian community so shares space with a pre school and a hula *halau* as well as other social services by and for Native Hawai'ians. Within each school, parents are expected to engage in their children's education in a variety of ways. At MCS students are expected to learn the Hawai'ian language along with their *keiki* (child). At all three schools, parents participate in parent-teacher conferences and attend performances of their children during the year. There is also a strong *kupuna* program that encourages grandparents and aunts and uncles to come to the school and work with students. They may teach Hawai'ian crafts, tell '*Olelo No'eau*, traditional stories, or perform more mundane tasks like serving lunch or accompanying students on excursions.

The curricula at these ethnocentric schools are often based on topics of particular relevance to Native Hawai'ian culture such as genealogy, navigation, and aquaculture. They also include instruction in traditional crafts and cultural practices. However, the pedagogy seems to reflect what Wells, Lopez, Scott and Holme (1999) characterize as progressive and student-centered pedagogy as distinct from factory-like "modern" public schools. All three schools report the use of project based, experiential, interdisciplinary curricula. They also use a variety of alternate assessment techniques and hands-on learning and performance-based tasks that are infused with technology. Two specific grants have provided state of the art computers and provide for gifted and talented education for all students. All of these characteristics are recommended practices for all students from all ethnic backgrounds.

Another distinction between the schools is the use of the Hawai'ian language. This varied considerably in the three schools in this sample. MCS relied on full immersion for k – 3rd grade students and introduced the formal study of English in 4th grade. KCS aimed to provide bilingual instruction and accommodate all Native Hawai'ian students. The language is important for the connection to the culture and deeper understanding of things Hawai'ian but not to the exclusion of English, the language of commerce and entrance into socio-economic self-sufficiency. HCS did not focus on the language for utilitarian reasons or language renaissance per se but used it as a connection to the community and connection to the past that would improve student perceptions of self in today's world.

Finally, these charter schools articulated a need to prepare students educationally for the future for different reasons. MCS wanted their k-6 students to be able to enter any middle school and be successful as speakers of Hawai'ian and agents of the culture. That has led them to reconsider the introduction of English instruction. KCS clearly expect its graduates to "perpetuate Hawai'i native culture, language and traditions into the next millennium"…and… "transform their neighborhoods into more sustainable communities, and agents for the preservation of Hawai'i's unique natural resources." They expect students to go to the community college or local university and return to their community to stop the out migration of successful Native Hawai'ians. HCS wants students "to sustain and develop the local, traditional community, natural environment and people. The children and school are resources that focus on community energy and pride. The Foundation that supports HCS is dedicated to making life in the community better and more prideful.

Although all three of the schools described in this study were located on the same island, they lie a significant distance from each other in communities with very different demographics, climates, and economic bases. However, Hawai'i's unique single-district educational system make them all part of the same structure. They are all painted both by their ethnocentricity and by their need to survive as charter schools in an essentially hostile environment. Indeed, it may be difficult to determine which plays a more central role in the formation of the character of the three schools: being a charter school or choosing an ethnocentric theme. What appears clear, however, is that organizations cannot develop on the strength of *what they are not*. Rather, even as efforts are made to break from educational practices which no longer serve the needs of our children, successful schools are those which stand *for* something, not *against* something.

The temptation to postpone judgment about the effectiveness of public ethnocentric charter schools in Hawai'i or the propriety of spending public funds on them is tempered by an appreciation of the disagreement over what constitutes "effectiveness" in this context. Proponents of objective normative evaluations of student learning could legitimately argue that the data aren't yet available. But our observations of the extreme satisfaction exhibited by all stakeholders in the ethnocentric charter schools we examined leads us to question the traditional criteria used to evaluate public schools. It is clear to us that these schools serve a purpose; they provide an education strongly preferred by its target client group which shows no obvious signs of being inferior to that provided by the over-burdened traditional system. If some normative evaluation is justified, it is nevertheless clear that it should not be the only criterion on which to assess the success of these schools.

Charter schools in Hawai'i, whether ethnocentric or not, are almost all associated with one or another special interest group. One, for instance, is clearly populated by children and grandchildren of the white "children of the sixties." Another was founded on the premise that nutrition (both its study and practice) is at the center of good learning. Still another relies heavily on the Waldorf approach. Nationally, charter schools can be found based upon a military/patriotic model or a Great Books (largely written by dead white males) curriculum. Each of these uses public funds for openly parochial purposes. It is our experience that these extremely diverse schools share essential characteristics: 1) their school communities are very satisfied and happy with them; and 2) they have had to overcome significant obstacles placed in their way by the traditional educational establishment.

In some aspects, the Hawai'ian ethnocentric charter schools we observed exemplify best practices that are almost universally acknowledged. They are small schools in a state which has the largest average school size in the country. They employ a self-selected group of teachers whose passion and enthusiasm lead them to endure significant hardship (tenure, retirement benefits, salary levels at charter schools are all issues in a state with universal public sector collective bargaining). High levels of parental involvement and community support for these charter schools are the envy of their traditional counterparts.

It may be easy to support the expenditure of public funds for ethnocentric charter schools in areas where ethnic minorities have traditionally been underserved. What is more problematic is contemplating what might happen if other special interest groups (ethnocentric or not) made similar educational arguments. Would one make the same supportive arguments in favor of an ethnocentric charter school in New York, for instance, seeking to connect students to their Italian roots, or a Chicago charter school conducted entirely in Polish? Are schools of choice schools of choice, or not?

We believe that the question is not what criteria to apply to distinguish schools of "good" choice from schools of "bad" choice. Rather, we should be looking at what this whole phenomenon presages for American public education. It seems likely to us that we are observing the opening rounds of a long term struggle between schools of choice and the traditional educational system. On one side are a growing number of individuals banded together into groups by their mutual interests and values who have stopped trying to fix the public school system in favor of struggling for the right to start their own. On the other side is a much larger group advocating the continuation of the current system and resisting change. The United States saw a similar phenomenon approximately one hundred fifty years ago.

We are witnessing a serious reassessment of some of American education's most cherished axioms. The inclusion of minorities loses its attractiveness when it is AGAINST THE WILL of those minorities. The maintenance of a free, appropriate public education loses its luster when clients challenge its appropriateness. Ultimately, we believe that public education is facing its own choices: lead, follow, or get out of the way.

Notes

1. New state legislation passed in May, 2002 (Senate Bill 2512, Hawai'i State Legislature), forbids lawsuits by Charter Schools against the Department of Education.

References

Asante, M. K. (2002). Imperatives of an Afrocentric curriculum. [On-line]. Available: http://twist.lib.uiowa.edu/tenthree/afroc.htm. (For additional information see:

http://www.asante.net/articles/index01.html)

Berv, J. (1998). Charter schools and the compromise of equity: An evaluation of Colorado's charter school legislation. (ERIC Document Reproduction Service No. ED 437 459)

Bolick, C. (1997). School choice, the law and the constitution: A primer for parents and reformers. The Heritage Foundation: Roe Backgrounder #1139. [On-line]. Available: http://www.heritage.org/library/categories/education/bg1139.html.

Buchanan, N. K. (1998) Education. . In S. P. Juvik & J. O. Juvik (Eds.) Atlas of Hawai'i (3rd ed.) (pp. 278-281). Honolulu, HI: University of Hawai'i Press.

Capozza, K. L. (1999). Education innovation: Indian alternative schools. *American Indian Report, 15* (7), 24-25.

Center of Urban and Public Affairs at the University of Texas - Arlington, Texas Center for Educational Research and Center for the Study of Educational Reform at the (2000). Texas open-enrollment charter schools: Evaluation 1998-99. [On-line]. Available: http://www.tea.state.tx.us/charter/eval99/index.html.

Cobb, C. D., & Glass, G. V. (1999). Ethnic segregation in Arizona charter schools. *Education Policy Analysis Archives, 7*(1), np. [On-line]. Available: http"//epaa.asu.edu/epaa/v7n1/.

Cobb, C. D., Glass, G. V. , & Crockett, C. (2000). The U.S. charter school movement and ethnic segregation. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April, 2000. (ERIC Document Reproduction Service No. ED 445 414)

Coffey, A. J. (2002) Magnets of many colors: Ethnocentric answers to education. [On-line]. Available: http://www.kip.jcomm.ohio-state.edu/ethnocentric_education.htm.

Columbia Electronic Encyclopedia (2000). (6th ed.). [On-line]. Available: http://www.encyclopedia.com/articles/04247.html.

Crockett, C. M. (1999). California charter schools: The issue of racial/ethnic segregation. Unpublished dissertation available at UMI ProQuest Digital Dissertations.

Dotts, C. K., & Sikkema, M. (1994). Challenging the status quo: Public education in Hawai'i 1840-1980. Honolulu, HI: Hawai'i Education Association.

Education Commission of the States (1999). Charter school equity. [On-line]. Available: http://www.ecs.org/clearinghouse/13/70/1370.htm.

Fusarelli, L. D. (2000). Texas: Charter schools and the struggle for equity. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April, 2000. (ERIC Document Reproduction Service No. ED 444 254)

Hawai'i State Department of Education (2002). School Community-Based Management Homepage. [On-line]. Available: http://www.Hawai'i.gov/scbm/scbm.html.

Hawai'i State Department of Education Planning, Budget, and Resource Development Office, (2001). School status and improvement reports. [On-line]. Available: http://doe.k12.hi.us/.

Hlebowitsh, P., & Tellez, K. (1997). *American education: Purposes and promise. Belmont*, CA: Wadsworth Publishing Company.

Kapono, E. M. (1998). Hawai'ian language renaissance. In S. P. Juvik & J. O. Juvik (Eds.) *Atlas of Hawai'i (3rd ed.)* (p. 199). Honolulu, HI: University of Hawai'i Press.

Kolbert, E. (2000, October 9). Unchartered territory: Is there money to be made off the failure of public education? *The New Yorker*, 34-41.

Langlas, C. M. (1998) History. In S. P. Juvik & J. O. Juvik (Eds.) *Atlas of Hawai'i (3rd ed.)* (pp. 169-182). Honolulu, HI: University of Hawai'i Press.

Merriam-Webster online dictionary. [On-line]. Available: http://www.m-w.com/cgi-bin/dictionary.

Moersch, C. (1994). Labs for learning: An experiential-base action model. Oregon: National Business Education

Alliance.

Nelson, B., Berman, P., Ericson, J., Kamprath, N., Perry, R., Silverman, D., & Solomon, D. (2000). The state of charter schools 2000 fourth year report. Accessed November 2000, Washington DC: Office of Educational Research and Improvement, U.S. Department of Education.

National Commission on Excellence in Education (1983). A nation at risk: The imperative of educational reform. Washington, DC: U.S. Department of Education.

National Study of School Evaluation (2002). [On-line]. Available: http://www.nsse.org.

Office of the Superintendent/Planning, Budget, and Resource Development Office, (2001). Hawai'i opinion poll on public education 2001. Honolulu, HI: Department of Education. [On-line]. Available: http://doe.k12.hi.us/.

Office of the Superintendent/Planning, Budget, and Resource Development Office, (2001). The superintendent's eleventh annual report on school performance and improvement in Hawai'i. Honolulu, HI: Department of Education. [On-line]. Available: http://doe.k12.hi.us/.

PRB/KIDS COUNT (2002). Children at Risk: State Trends 1990-2000, Special Report. [On-line]. Available: http://www.aecf.org/kidscount/c2ss/.

Peterson, P. E. (1998). Top ten questions asked about school choice. In D. Ravich (Ed.), BPEP Conference on the State of Urban Education in America. (pp. 57-62). The Brookings Institute, Washington DC: The Brookings Institute, Brown Center on Education Policy.

Public Sector Consultants Inc., & MAXIMUS Inc. (2001). Michigan Charter School Report. [On-line]. Available: http://www.voyager.net/psc/charter/index.html.

Ravitch, D. (1990). Multiculturalism: E Pluribus Plures. *The Key Reporter*, 56 (1), 1-4.

Rebus Planning Associates, Inc. (1994). *The work sampling system* (3rd edition). Ann Arbor, MI: Author.

Rees, N. S. (2000). School choice 2000 annual report, The Heritage Foundation: Roe Backgrounder # 1354. [On-line]. Available: http://www.heritage.org/library/categories/education/bg1354.html.

Sarason, S. B. (1998). *Charter schools: Another flawed educational reform?* NY: Teachers College Press.

Shokraii, N. (1996). Free at last: Black Americans sign up for school choice. *Policy Reviews, The Journal of American Citizenship*. November/December #80. [On-line]. Available: http://www.policyreview.org/nov96/backup/shokraii.html.

Schmitt, R. C. (1998).Population. . In S. P. Juvik & J. O. Juvik (Eds.) *Atlas of Hawai'i (3rd ed.)* (pp. 183-197). Honolulu, HI: University of Hawai'i Press.

Tyack, D. B. (1974). *The one best system*. Cambridge, MA: Harvard University Press.

Wells, A.S., Artiles, L., Carnochan, S., Cooper, C.W., Grutzik, C. Holme, J., Lopez, A., Scott, J., Slayton, J., & Vasudeva, A. (1998). Beyond the rhetoric of charter school reform: A study of ten California school districts. UCLA.

Wells, A. S., Lopez, A., Scott, J, & Holme, J. J. (1999). Charter schools as postmodern paradox: Rethinking social stratification in an age of deregulated school choice. *Harvard Educational Review, 69* (2), np. [On-line]. Available: http://www.edreview.org/harvard99/1999/su99/s99wells.htm.

Whitty, G. (1997). Creating quasi-markets in education: A review of recent research on parental choice and school autonomy in three countries. In M. W. Apple (Ed.), *Review of Research in Education* (pp. 3-47). Washington DC: American Educational Research Association.

About the Authors

Nina K. Buchanan is an Educational Psychologist and Professor of Education at the University of Hawai'i at Hilo and is Co-Director the UH Charter School Resource Center. She has co-edited (with John F. Feldhusen) *Conducting Research and Evaluation in Gifted Education: A Handbook of Methods and Applications*. Her latest work has been in talent development through innovative high school programs. She is a contributing editor of Roeper Review and serves on the Local School Boards of the West Hawai'i Explorations Academy and Wai Ola Public Charter Schools.

Email: ninab@hawaii.edu

Robert A. Fox is a Professor of Physics in the Department of Physics and Astronomy, University of Hawai'i at Hilo where he also serves as Co-Director of the University of Hawai'i Charter School Resource Center. He has served as an elected member of the Hawai'i State Board of Education. Current research interests include the affect of collective bargaining on charter schools and the struggle between charter schools and the traditional educational establishment.

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

This article has been retrieved 4944 times since February 28, 2003

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Education Policy Analysis Archives

Volume 11 Number 9 **February 28, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the **EDUCATION POLICY ANALYSIS ARCHIVES**. Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Creating a System of Accountability: The Impact of Instructional Assessment on Elementary Children's Achievement Test Scores

Samuel J. Meisels
Erikson Institute

Sally Atkins-Burnett
University of Michigan

Yange Xue
Julie Nicholson
Palo Alto, CA

Donna DiPrima Bickel
University of Pittsburgh

Seung-Hee Son
University of Michigan

Citation: Meisels, S. J., Atkins-Burnett, S., Xue, Y., Nicholson, J., Bickel, D. D., and Son, S-H. (2003, February 28). Creating a system of accountability: The impact of instructional assessment on elementary children's achievement test scores, *Education Policy Analysis Archives, 11*(9). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n9/.

Abstract

This study examined the trajectory of change in scores on the Iowa Tests of Basic Skills (ITBS) of low-income, urban, third and fourth graders who had been enrolled in classrooms where the Work Sampling System (WSS), a curriculum-embedded performance assessment, was used for at least three years. The ITBS scores of children exposed to WSS were compared with those of students in a group of non-WSS contrast schools that were matched by race, income, mobility, school size, and number of parents in the home and to a comparison group of all other students in the school district.

Results indicated that students who were in WSS classrooms displayed growth in reading from one year to the next that far exceeded the demographically matched contrast group as well as the average change shown by all other students in the district. Children in WSS classrooms made greater gains in math than children in the other two groups, although the results were only marginally significant when compared with gains by the matched contrast group. The discussion concerns the complementarity of performance-based and normative tests in systems of accountability and the potential value of using a curriculum-embedded assessment to enhance teaching, improve learning, and increase scores on conventional accountability examinations.

More group-administered achievement testing is taking place in states and local school districts than ever before (Achieve, Inc., 2002). According to a survey published in *Education Week* (Olson, 2001), every state has adopted mandatory tests at one or more grades in elementary, middle, and high school, and 49 states have linked their academic standards to these tests. Moreover, legislation mandating annual testing in reading and mathematics for all children in grades 3 – 8 was recently enacted by Congress.

The tests in use nationwide are standards-based, but primarily norm-referenced; only 10 states report use of supplementary criterion-referenced tests (Olson, 2001). The principal purpose of these tests is to ascertain the current status of student achievement, rather than to identify students in need of intervention or to determine appropriate instructional strategies. Only seven states provide extra funding for low-performing schools and just nine states allocate funds for remediation of failing students. By the close of the year 2000, 18 states had made graduation contingent on student test performance and an additional five states were about to begin administering exit exams. As many as 27 states could withhold diplomas to students who fail to pass state accountability examinations by 2003 (Olson, 2001).

In addition to the increased prevalence of such tests and the escalation of consequences or "stakes" associated with them, the most notable change in the assessments themselves concerns their alignment with curriculum standards. Never before in U. S. history have we witnessed such an explosion of attention to standards. The minimum competency testing movement of the 1980s that was inaugurated by the *Nation At Risk* report (National Commission on Excellence in Education, 1983) has been overtaken by a focus on high standards of achievement inspired by the nation's embrace of the national educational goals of the 1990s.

Ironically, one result of the standards-based reform movement has been a heightened emphasis on high-stakes, group-administered, decontextualized testing practices. This is due to the linkage between standards-based reform and the national political drive for accountability (Kohn, 2000). As Popham (2000) points out, these tests lend themselves to the "score-boosting game" in which educators devote most of their energy to raising students' scores on conventional achievement tests. Rather than trying to improve student performance by enhancing instruction, this approach views high stakes testing as a policy tool to leverage learning (Firestone & Mayorowetz, 2000). In short, funding, availability of other resources, and state and local prestige are all devoted to improving student test scores by changing the curriculum to match more closely the items or content standards of the assessments. This practice of "curriculum alignment" has the consequence of increasing test scores for some students, but often leaving general knowledge and mastery of curriculum domains virtually untouched (Amrein & Berliner, 2002; Corbett & Wilson, 1991; Haney, 2000; McNeil, 2001). High-stakes testing has been shown to impair the validity of the tests themselves as the test-taking experience becomes less a sampling of students' adaptive skills and higher-order thinking and more of an exercise in rote memory and mastery of basic skills (Orfield & Kornhaber, 2001). In a study of 18 states with high-stakes testing programs, the learning levels of the students in all but one of the states were at the same level as before the testing policies were implemented (Amrein & Berliner, 2002).

The alternative to teaching to the test, or measurement-driven instruction, is to transform instruction, but to do so in such a way that the standards that are intended to serve as the basis of the tests inform instructional decisions and are incorporated into new forms of assessment. Curriculum-embedded performance assessments (see Baron & Wolf, 1996; Darling-Hammond, 1992) represent an instructional-driven measurement in which students' actual classroom performance is evaluated in terms of standards-infused criteria. These criteria in turn suggest next steps in curriculum development which are consistent with advancing progress toward attainment of the defined standard. It is reasonable to assume that as students' learning improves, so will their scores on accountability examinations.

Unfortunately, the research literature contains few studies of the impact of curriculum-embedded performance assessments on group-administered achievement test scores (Borko, Flory, & Cumbo, 1993; Falk & Darling-Hammond, 1993). Rather, most studies of the impact of testing provide data about conventional results-driven accountability tests—tests that rely on public reporting of performance data and utilization of these data for reward or sanction. This approach to assessment (which can also be called "norm-referenced accountability") emphasizes the use of test data for instrumental purposes—purposes external to the classroom—rather than direct application of the data to improve educational practice.

In this paper we describe an alternative to typical conceptions of accountability. Instead of relying on either norm-referenced or performance-based assessments in isolation, we suggest a complementary approach that incorporates both types of assessment. In short, consistent with recent federal and state initiatives, we suggest the addition of performance assessments to conventional norm-referenced testing.

This paper investigates whether students who are enrolled in classrooms in which a curriculum-embedded performance assessment is in use will show greater gains on a conventional test used for accountability than students who have not had exposure to the performance assessment. The research question this study poses is the following: Can ongoing, focused instructional assessments influence performance on group-administered achievement tests? A corollary to this question concerns whether instructional and high-stakes assessments can be linked to create an accountability system that relies on both classroom- and test-based information about student achievement.

Method

Procedures

This study is part of a larger investigation that evaluated the validity of a curriculum-embedded performance assessment—the Work Sampling System (WSS; Meisels, Jablon, Marsden, Dichtelmiller, & Dorfman, 1994, 2001)—and its influences on teacher practices and children's achievement in the Pittsburgh Public Schools (PPS). The overall study included data from teachers, parents, and children. This paper focuses on the impact of WSS on the trajectory of children's change in scores on a group-administered achievement test (the Iowa Tests of Basic Skills [ITBS]) from grade 3 to grade 4. Related studies focus on other aspects of the validity of WSS (Meisels, Bickel, Nicholson, Xue, & Atkins-Burnett, 2001), parental reactions to WSS (Meisels, Xue, Bickel, Nicholson, & Atkins-Burnett, 2001), and teachers' views of the consequences of using WSS (Nicholson, 2000).

The Work Sampling System (Meisels, 1997; Meisels et al., 1994; 2001) is a curriculum-embedded performance assessment designed for children from preschool through grade 5. WSS is comprised of developmental guidelines and checklists, portfolios, and summary reports. It uses teachers' perceptions of their students in actual classroom situations as the data of assessment while simultaneously informing, expanding, and structuring those perceptions. It involves students and parents in the learning and assessment process and it makes possible a systematic documentation of what children are learning and how teachers are teaching. This approach to performance assessment allows teachers the opportunity to learn about children's processes of learning by documenting children's interactions with materials, adults, and peers in the classroom environment and using this documentation as the basis for evaluating children's achievements and planning future educational interventions through comparisons with standards-based guidelines. Evidence of the reliability and validity of Work Sampling is available in Meisels, Bickel, Nicholson, Xue, and Atkins-Burnett (2001) and Meisels, Liaw, Dorfman, and Nelson (1995). Further descriptions of WSS are found in Meisels (1996, 1997) and Meisels, Dorfman, and Steele (1995).

The Iowa Test of Basic Skills (ITBS; University of Iowa and Riverside Publishing, 1994) is a group-administered achievement test designed to monitor year-to-year achievement differences in students from K – Grade 12. Norming for the 1993 edition (Form K) was completed on 136,934 individuals. KR-20 internal consistency ratings are reported at >.84 for all reading and mathematics subtests. The third and fourth graders in this study were administered the reading comprehension and the vocabulary subtests of the ITBS Survey Battery (Levels 7 - 11). The Reading total score includes both the comprehension and vocabulary scores. The Mathematics total score includes items tapping computation, estimation, calculation, problem solving, and data interpretation (Levels 7 - 11). Developmental Standard Scores (DSS) were calculated using the raw score/DSS conversion tables provided by ITBS (Hoover et al., 1993). The DSS scales allow a comparison of both status and growth even when children take different level tests. The mathematics computation score is combined with the total score to determine the DSS for mathematics plus computation.

Design

This report describes the results of a natural experiment in which two groups of demographically matched students who were administered the ITBS were compared with one another and with all other students in their grade levels in the PPS who were also administered the ITBS. One of the target groups was composed of students who had been exposed to WSS for three years prior to being administered the ITBS; all other students had no experience with WSS. The ITBS was administered to all children by their teachers in the spring of 1997 (end of third grade) and the spring of 1998 (end of fourth grade). A total of 2708 students received the ITBS reading assessments in both 1997 and 1998 and 2564 students took the ITBS math assessments in both years. Data were coded and reported by school district personnel.

A longitudinal design was selected because the schools using WSS were among the lowest-performing schools on the ITBS in the district. It was evident to us that comparisons of absolute scores at the end of third or fourth grade would only confirm that the children in these low-scoring schools were still low scoring, despite potential improvements in comparison with children from the same schools but from different age cohorts. The longitudinal design focuses on the trajectory of change from third to fourth grade as a way of capturing growth over time *within* students. This intra-individual use of normative data enabled us to examine the relative change in student achievement without regard for comparisons of absolute differences in student scores.

WSS was adopted as part of the District's restructuring effort, but it was not the sole innovation taking place in the district. New reading, mathematics, and social studies curricula were introduced at the same time as WSS, however, not all schools and classrooms in the district implemented all of these practices. Since information was unavailable about which schools implemented which instructional innovations, it is not possible to test alternative explanations for our results by isolating WSS versus any of these interventions.

Sample

At the time this study took place WSS had been used in a sample of the Pittsburgh Schools for three years. All teachers in the WSS schools were voluntary participants. Selection criteria for participation of teachers in this study included use of WSS for at least two years and a determination of the fidelity of implementation of participating teachers. This was ensured by a review by external examiners conducted in the spring of 1996 of portfolios and of the teachers' 1996–97 WSS materials by the research staff. These selection criteria limit the generalizability of our results but provide a test of the full implementation of WSS.

The longitudinal data consist of 96 third grade students in the WSS schools, 116 students in the non-WSS comparison schools, and 2922 students enrolled in all other PPS Grade 3 and 4 classrooms in 1996–98. For students in WSS schools, 71% were African-American and 90% received free or reduced lunch (see Table 1). There were more girls (58%) than boys in the sample. To form the comparison group, classrooms were chosen that matched those in the WSS schools as closely as possible on race, income, mobility, school size, and number of parents in the home. In other PPS schools, 70% of the students were African-American and 87% received free or reduced lunch (see Table 1).

Table 1
Demographic characteristics of WSS schools, comparison schools, and other PPS

Group	School Size (N)	F/R Lunch (%)	Other Parents in Household (%)	Mobility (%)	African-American (%)
WSS	400.4	90.0	70.6	9.8	70.6
Comparison	298.8	89.8	74.0	9.4	75.2
PPS Other	311.2	87.0	70.8	9.4	71.5

WSS = Work Sampling System
PPS = Pittsburgh Public Schools

Analysis

Comparisons of mean change in reading and math scores on the ITBS from Grades 3 to 4, as well as regression analyses, were conducted in order to study the average change in test scores from one year to the next among the three groups. As noted, analysis of previous school district results showed that the WSS schools scored at or near the bottom of the district's ITBS test scores. By controlling for initial ability we were able to study the trajectory of change in students' ITBS scores from one year to the next in order to determine if differential change on the ITBS took place despite the low levels of initial competence on this assessment shown by students in the WSS schools.

Our analytic approach began with a comparison of the change in ITBS scores of the WSS third graders in 1997 - 1998 as compared with a comparison group matched on key demographic characteristics and all other PPS third graders in that cohort. Differences between the 1997 and 1998 ITBS Developmental Standard Scores (DSS) were computed to create DSS change scores, which we used as one indicator of academic growth. In order to examine whether WSS had a differential impact on high or low achievers, the means of the 1997 DSS scores for the PPS district were used to divide all the PPS students into above average and below average groups. The gains for high and low achieving students within each group (WSS, comparison, and all other PPS) were then compared.

Next, three-step hierarchical regressions were performed to examine the relative effect of participating in WSS on the students' change in performance from third to fourth grade. Because change scores may be particularly sensitive to problems with floor and ceiling, we used a covariance model with the Grade 4 score as the outcome, controlling for initial ability and the level of form administered. In the first step, children's 1997 DSSs were entered to control for differences in initial ability levels. Typically, fourth graders take a Level 10 form, however, PPS allows for administration of out-of-level tests. Although the publishers equated the forms, and the level of administration should have had minimal effect on the DSS, we entered the level form into the regression to ensure that results were not biased due to differences in forms or variations in administration. Children taking a below-grade form might be more apt to reach the ceiling of the measure and receive an inflated estimated ability. Only three children in math and two children in reading took a form above the grade level; we excluded these cases from analyses. For the remaining students we created a dummy variable that was entered in the second step of the regression to indicate if the student received a below-grade level form for the Grade 4 administration. (The percentage of students in below-grade level forms was similar in the WSS and all other PPS groups—10 and 11% respectively. Only 6% of students in the comparison schools took below-levels forms.) Finally, students' group membership (comparison and all other PPS with WSS as the referent group) was entered into the regression model. Since dependent variables (1998 DSS scores) and the children's initial scores (1997 DSS scores) were standardized. The regression coefficients can be interpreted as effect sizes.

Missing Data

In order to study the impact of missing data on our conclusions, the means of the students with missing subtests were compared by group membership. In 1997, 21.9% of the WSS students were missing the reading total score; 15.6% were missing the 1998 reading total score. The comparison group was missing 13.8% of its reading total scores in 1997 and 17.2% in 1998. Among the other PPS students, 3.7% were missing the 1997 reading total score and 11.2% were missing the 1998 reading total score. Due to a missing score in either of the years, 30.2% of the WSS students, 17.2% of the comparison students, and 13.4% of other PPS students were excluded from the reading change score analysis. In mathematics, 7.3% of the WSS group was missing the 1997 math total score compared with 13.8% of the comparison group and 5.7% of the other PPS. For the 1998 math total score 18.8% of the WSS group, 17.2% of the comparison group, and 16.1% of the other PPS students had missing data. Thus, 18.8% of the WSS group, 22.4% of the comparison group, and 19% of the other PPS group were not included in the analysis of the mathematics change score.

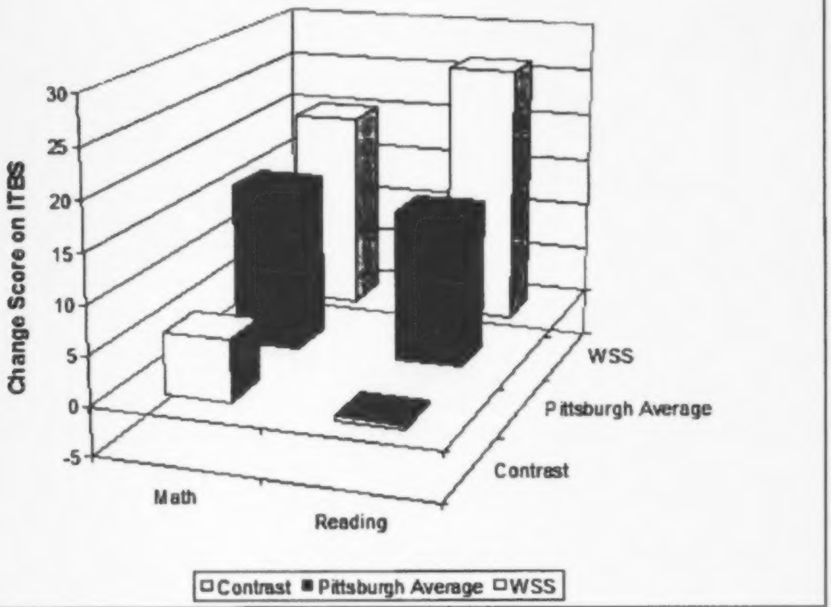
For the WSS group, we compared students' missing scores for one of those years with students who had scores for both years in terms of gender, ethnicity, age, SES, and Woodcock Johnson Psychoeducational Battery-Revised (WJ-R) broad reading, broad math, and broad writing scores in third grade. (The WJ-R was available only for the WSS group.) The results showed no significant differences across these variables except that fewer students who were missing at least one mathematics score received free or reduced lunch (58% vs. 92%, $p < .01$). There were no gender or race differences between missing and non-missing students for the comparison group.

The means for the ITBS subtests that were taken by the missing students were compared to the means of the other students in that group. For the WSS group, the means on the mathematics and reading total scores were not significantly different between the missing and non-missing groups. However, the 1997 reading DSS score was significantly lower for the comparison group students who were missing the reading subtests in one of those years (171 vs. 197, $p < .05$), but there were no significant differences in mathematics total score between the missing and non-missing students for the comparison group. For the other PPS students, the reading total scores in both 1997 and 1998 were significantly lower among the missing group (173 vs. 181 in 1997, $p < .001$; 186 vs. 197 in 1998, $p < .01$), but the means on the mathematics total score were not significantly different for the missing and non-missing groups. Thus, the effects of the missing data on our findings were relatively limited.

Results

In this analysis we compare ITBS scores of third and fourth grade WSS and non-WSS students (both the matched comparison group and all other PPS students) in order to determine if differential achievement on this outcome is associated with participation in WSS classrooms. We begin by comparing the change in mean DSS in reading and math among the WSS students, the comparison group, and the remainder of the PPS students (see Figure 1). The mean change scores of the WSS group (27 and 20 points for reading and math, respectively) are substantially greater than those of the other groups (0 and 6 for the comparison and 15 and 17 points for all other PPS in reading and math, respectively). The differences in the groups are particularly strong in reading, with a mean change on the reading total DSS that is more than 11 points greater for the WSS students when compared with the other PPS students ($t = 4.33$; $p < .001$). The moderately large effect size of .68 indicates meaningful as well as significant differences in reading change scores (Cohen, 1988). When considered in relation to the comparison group, the discrepancy is even greater. The WSS group mean change score is more than 27 points higher than the comparison group's mean change score ($t = 8.86$; $p < .001$). The unusually large effect size ($d = 1.60$) indicates strong differences in the sample in measured change in reading.

Figure 1. Mean Change Scores on ITBS for Third-Fourth Grader



In math, results follow a similar pattern, although they are not as dramatic. The mean change score of the WSS students is greater than that of the other PPS students by more than three points, a marginally significant finding ($t = 1.89$; $p = .059$). The examination of the effect size ($d = .20$) suggests a small but nontrivial effect (Cohen, 1988). The mean change of DSS math score in WSS students is almost 14 points higher than that of the comparison group ($t = 4.88$; $p < .001$), indicating a large effect ($d = .76$).

To investigate whether WSS had a differential effect on high and low achieving students, we used a segmentation analysis. The mean 1997 DSS ratings of the entire sample were used to divide all students into above and below average groups (see Table 2). Before comparing differential changes of score in these groups, we examined the initial scores of the 1997 DSS. As expected, the initial scores of the WSS students were lower than the comparable group (above and below average) in the other samples (comparison and other PPS), although the mean differences were relatively small (2-4 point differences) and not highly significant, with the exception of the above average group of comparison students in reading. The mean of the 1997 reading DSS of the above average students differed by more than one SD between WSS and the comparison group and the above average performers represented a greater percentage of the comparison sample than was the case in the WSS sample.

Table 2
Mean of 1997 DSS in above and below average performers

Subtest	WSS			Comparison			Other PPS			WSS vs. Comparison	WSS vs. Other PPS
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*t*	*t*
Reading total											
Above Average	191.40	11.05	10	205.36	17.15	77	199.91	15.90	1295	-3.49**	-1.69
Below Average	160.69	13.17	65	164.81	13.87	31	164.12	12.13	1519	-1.41	-2.22*
Math total											
Above Average	197.58	14.56	24	202.31	15.48	84	199.30	13.31	1534	-1.34	-.63
Below Average	163.57	9.95	65	168.61	9.99	23	167.19	9.42	1377	-2.09*	-3.17**

Note. *t* = t-score, DSS = Developmental Standard Score
* $p<.05$, ** $p<.01$

With one exception, above and below average WSS students in reading and math made gains that were greater than the comparison group and the other PPS students (see Table 3). With small differences in initial ability in all but one area, the effect sizes of the differences in change scores were all moderate to high in reading. In mathematics the results were more equivocal. Among below average math achievers, WSS had significantly higher change scores than other PPS students (t = 2.14; $p < .05$; $d = .29$). The change scores of low performing PPS students in math were greater than the change scores of either the WSS or the comparison group of low performers. However, these results were not significant and the effect size was negligible ($d = .05$).

Table 3
Mean of DSS change in above and below average performers (SD)

Subtest	WSS			Comparison			Other PPS			*WSS vs Comparison*		*WSS vs Other PPS*	
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*t*	*d*	*t*	*d*
Reading total													
High	26.56	20.44	9	-4.01	17.42	72	12.29	16.78	1193	4.87***	1.77	2.54*	0.82
Low	26.67	21.67	58	10.50	12.71	24	17.93	15.44	1337	3.41***	1.02	3.04**	0.55
t (high vs. low)	.02			3.76***			8.76***						
d (high vs. low)	.00			.83			.35						
Math total													
High	8.32	23.31	19	2.77	6.26	70	12.91	23.31	1138	.94	.24	-.25	.05
Low	24.25	17.18	59	18.85	15.49	20	20.04	16.79	1230	1.21	.32	2.14*	.29
t (high vs. low)	3.21**			3.94***			10.31***						
d (high vs. low)	.80			.83			.39						

Note. *t* = t score, *d* = standardized estimates of effect size
*$p < .05$, ** $p < .01$, *** $p < .001$

The publishers of the ITBS contend that children with above average ability will show greater gains than other children from year to year. However, in each of the three samples, lower performing students made greater gains than higher performers. The difference in the mean reading change score between high and low performing students was greater in the comparison group (14.51) and the PPS group (5.64) than in the WSS group (0.12). The WSS group did not show substantial differences in reading score change between high and low performers (t = .02; p = N.S), although there were significant differences between high and low performing students in the comparison group (t = 3.76; $p < .001$; $d = .83$) and the other PPS group (t = 8.76; $p < .001$; $d = .35$). This suggests that high and low performing students profit equally from WSS in reading. In mathematics, the other PPS group demonstrated the least difference in change scores between high and low performing students (7.13) compared with the WSS group (15.94) and the comparison group (16.08). The difference in change scores between high and low performing students was significant in WSS (t = 3.21; $p < .01$; $d = .80$) and the comparison group (t = 3.94; $p <. 001$; $d = .83$), as well as the other PPS groups (t = 10.31; $p <. 001$, $d = .39$).

To examine whether group membership accounts for differences in achievement growth after controlling for differences in both initial achievement (i.e., 1997 DSS) and the level form taken in 1998, three step regressions were performed .The initial achievement and the 1998 level form (below-grade level form = 1, on-grade level form = 0) were entered in the first and the second step, respectively. Then, group membership was entered in the third step. All three models were statistically significant (see Table 4). Initial ability, level of form administered, and membership in WSS all predicted students' fourth grade reading DSS (see Table 4). Children who took on-grade level forms had fourth grade reading DSS more than one-half SD higher than those who took below-grade level forms, even after controlling for initial ability ($p < .001$). The children who were in WSS schools had higher fourth grade scores in reading by .17 SDs when compared with the other PPS students, after controlling for the effect of initial achievement and the level form taken in 1998 ($p < .05$). The average fourth grade score for the WSS group was .60 SDs greater than the average score for the comparison group after controlling for initial achievement and 1998 level form ($p < .001$)

Table 4
Covariance models of Grade 4 achievement

	Grade 4 DSS Reading Total (N=2,772)			Grade 4 DSS Mathematics Total (N=2828)		
Variable	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3
Initial ability[a]	.74 ***	.67***	.68***	.56***	.56***	.56***
1998 Level Form (below-grade level)		-.55***	-.54***		-.10*	-.10*
Other PPS group			-.17*			-.06
Comparison group			-.60***			-.18
R^2	.55***	.58***	.58***	.32***	.32***	.32***
R^2 change		.03***	.01***		.00*	.00

Note. All regression coefficients are shown in effect size using standardized scores.
DSS = Developmental Standard Score
* $p < .05$, *** $p < .001$
[a]Initial ability is the third grade DSS on the respective measures (reading and mathematics)

The covariance model for the mathematics total fourth grade DSS showed a somewhat different pattern. Group membership was not a significant predictor of the fourth grade mathematics DSS after controlling for initial ability and level form taken. Students who had higher mathematics DSS on the 1997 ITBS ($p < .001$) and took an on-grade level form in 1998 ($p < .05$) had higher fourth grade scores than other students. However, the variance explained by this model was unusually low for a covariance model that uses the identical assessment to control for initial ability ($R^2 = .32$).

Discussion

This paper examined whether students enrolled in classrooms that use a curriculum-embedded performance assessment will show greater gains on a conventional test of academic accountability than students without exposure to such a performance assessment. Our results indicate that students who were in WSS classrooms display growth in reading from one year to the next that far outstrips a demographically matched comparison group and that also exceeds the average change shown by all other students in the district. Further, by examining the results of above and below average students separately, we were able to demonstrate that the impact of the curriculum-embedded performance assessment is not limited solely to those who start with either low or high skills. Rather, the impact appears to be across the board with high and low performing students making comparable gains in reading. This analysis effectively dismisses the objection that these results are attributable to regression to the mean, since students with higher scores in *both* groups performed better than low scoring students in either group. The three-step regressions further demonstrate that participation in WSS classrooms accounts for these differences even after taking into account initial achievement ranking and level of test form administered.

The pattern of change is similar though not as strong in mathematics, although it does not appear to benefit high and low performing students equally. Students who were in the WSS group had higher mathematics scores after controlling for initial ability and level of form administered, but these findings were not statistically significant with this size sample (the effect size with the other PPS group was .18). In addition, the segmentation analysis indicated that the mean change score was lower for high performing students in all groups than for below-average students.

Examination of the construction of the DSS on this edition of the ITBS provides a clue to potential reasons for this finding. In this version of the ITBS, the mathematics DSS appears to be heavily weighted by number concepts and operations and, in particular, by computation (procedural knowledge of operations). In contrast, national standards in effect at the time (NCTM, 1989) and WSS (which is based on these and other standards) address multiple strands of mathematical thinking, including geometry, measurement and spatial sense, data analysis, statistics and probability, patterns, functions, and algebra, as well as number concepts and operations. The findings in mathematics may be a reflection of a mismatch between the broad standards WSS reflects and the test specifications of this edition of the ITBS. In addition, the covariance model for mathematics suggests that there is a problem with this measure of mathematics skills. Only 32% of the variance in the fourth grade scores could be explained by the third grade scores.

The findings on the reading assessment are robust and pervasive and make an important contribution to discussions of accountability. Whether looking at high or low performing students, examining gain scores or using a covariance model, students enrolled in WSS classrooms made greater gains in reading than students who did not have this exposure. These findings, as well as those for math, though not definitive because of our inability to disentangle the impact of WSS from the other innovations co-occurring in the district, suggest a new way to approach accountability testing. For too long, accountability examinations have been assumed to be of a particular kind with an unambiguous focus: normative assessments intended to rank students numerically and compare them to the performance of a specified group. As noted at the outset of this paper, it is likely that the U.S. is spending more money on tests at this time than at any previous point in its history. However, children are not faring well on these assessments. Media reports of large numbers of failures in numerous states are interpreted either in terms of students' lack of skills or teachers' inability to align curricula to the standards that are used to design high stakes tests (Manzo, 2001).

However, another explanation is also possible, and this perspective provides a link to our second research question: Is it possible to design an accountability system that relies on both classroom- and test-based information about student achievement? The alternative view of the accountability debate is that we not only need high standards and tests that reflect these standards, we also need curricula that will enable students to be successful on these assessments—but that are not simply instances of measurement-driven instruction.. By implementing an instructional assessment such as Work Sampling, teachers obtain information about their students on a continuous basis across multiple curriculum domains and from several assessment sources. They compare student performance with standards-based guidelines. They collect multiple sources of information from checklists, portfolios, and student and parent reports. They engage in curriculum analysis in order to evaluate artifacts included in portfolios. And they participate in processes of planning, review, and analysis with their colleagues. Students also have a meaningful role in the assessment process and thus become active participants in the evaluation process by becoming more familiar with the standards and how to progress towards those standards. This appears to enhance teaching and improve learning.

Perhaps the most important lesson that can be garnered from this study is that accountability should not be viewed as a test, but as a *system*. When well-constructed, normative assessments of accountability are linked to well-designed, curriculum-embedded instructional assessments, children perform better on accountability exams, but they do this not because instruction has been narrowed to the specific content of the test. They do better on the high stakes tests because instruction can be targeted to the skills and needs of the learner using standards-based information the teacher gains from ongoing assessment and shares with the learner. "Will this be on the test?" ceases to be the question that drives learning. Instead, "What should I learn next?" becomes the focus.

When accountability is seen as a system that incorporates both instructional assessment and on-demand tests, both teaching and learning can be affected positively. Moreover, this methodology provides policy makers with clear documentation not only of summative accomplishments, but also of the process of teaching and learning. The approach described in this study places emphasis where it belongs: on teaching and learning, rather than on testing. And it does so without sacrificing either the student or the teacher on the altar of accountability.

Acknowledgement

We wish to thank Jack Garrow for assisting us in obtaining and interpreting school district data. We are also deeply grateful to the principals, teachers, parents, and children who participated in this study, and to the staff and administrators of the Pittsburgh Public Schools. This study was supported by a grant from the School Restructuring Evaluation Project, University of Pittsburgh, the Heinz Endowments, and the Grable and Mellon Foundations. The views expressed in this paper are those of the authors and do not necessarily represent the positions of these organizations.

References

Achieve, Inc. (2002). *Achieve + McRel Standards Database*. http://frodo.mindseye.com/achieve/achieved.nsf

Amrein, A. L. & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved 4/1/02 from http://epaa.asu.edu/epaa/v10n18/.

Baron, J. B. & Wolf, D. P. (Eds.), *Performance-based student assessment: Challenges and possibilities* (95th Yearbook of the Society for the Study of Education, pp. 166-191). Chicago, IL: Society for the Study of Education.

Borko, H., Flory, M., & Cumbo, K. (October, 1993). *Teachers' ideas and practices about assessment and instruction. A case study of the effects of alternative assessment in instruction, student learning, and accountability practice.* CSE Technical Report 366. Los Angeles: Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Wiley

Corbett, H. D. & Wilson, B. L. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex.

Darling-Hammond, L. (1992). *Standards of practice for learner-centered schools*. New York: Columbia University, Teachers College, National Center for Restructuring Education, Schools, and Teaching.

Falk, B., & Darling-Hammond, L. (March, 1993). *The primary language record at P.S. 261: How assessment transforms teaching and learning*. New York: National Center for Restructuring Education, Schools, and Teaching.

Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., Dunbar, S. B., Oberley, K. R., Cantor, N. K., Bray, J. C., Lewis, J. C., & Qualls-Payne, A. L. (1993), *Iowa Tests of Basic Skills Norms and scores conversions Form K Survey Battery*. Chicago: Riverside.

Jennings, J. (1998). *Why national standards and tests? Politics and the quest for better schools*. Thousand Oaks, CA: Sage Publications, Inc.

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

Manzo, K. K. (2001, June 20). More than half of California 9th graders flunk exit exam. *Education Week*, p. 19.

Meisels, S. (1996). Performance in context: Assessing children's achievement at the outset of school. In A. Sameroff and M. Haith (Eds.), *The five to seven year shift: The age of reason and responsibility* (pp. 407-431). Chicago: The University of Chicago Press.

Meisels, S. (1997). Using Work Sampling in authentic assessments. *Educational Leadership, 54* (4), 60-65.

Meisels, S., Dorfman, A., & Steele, D. (1994). Equity and excellence in group-administered and performance-based assessments. In M. Nettles and A. Nettles (Eds.), *Equity in educational assessment and testing* (pp. 195-211). Boston: Kluwer.

Meisels, S.J., Jablon, J., Marsden, D.B., Dichtelmiller, M.L., & Dorfman, A. (1994). *The Work Sampling System* (3d ed.). Ann Arbor, MI: Rebus Inc.

Meisels, S.J., Jablon, J., Marsden, D.B., Dichtelmiller, M.L., & Dorfman, A. (2001). *The Work Sampling System* (4th ed.). Ann Arbor, MI: Rebus Inc.

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in Kindergarten-Grade 3. *American Educational Research Journal, 38* (1), 73 – 95.

Meisels, S.J., Liaw, F-r, Dorfman, A., Nelson, R. F. (1995). The Work Sampling System: Reliability and validity of a performance assessment for young children. *Early Childhood Research Quarterly, 10*, 277-296.

Meisels, S. J., Xue, Y., Bickel, D. D., Nicholson, J., & Atkins-Burnett, S. (2001) Parental reactions to authentic performance assessment. *Educational Assessment, 7* (1), 61-85.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics, grades K-6*. Reston, VA: Author.

Nicholson, J. M. (2000). *Examining aspects of consequential validity in a curriculum-embedded performance assessment*. Doctoral dissertation, Ann Arbor, MI: University of Michigan.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: A report to the nation and Secretary of Education, United States Department of Education*. Washington, D.C.: The Commission [Supt. Of Docs., U.S. G.P.O. distributor].

Orfield, G. & Kornhaber, M. L. (Eds.) (2001). *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Foundation Press.

Popham, W. J. (2000). *Testing! Testing! What every parent should know about school tests*. Boston: Allyn & Bacon.

Olson, L. (2001, January 11). Finding the right mix. *Education Week, Quality Counts Special Issue*, p. 12 - 20.

University of Iowa and Riverside Publishing Co. (1994). *Riverside 2000 Integrated Assessment Program: Technical Summary I.* Chicago: The Riverside Publishing Company.

About the Authors

Samuel J. Meisels
Erikson Institute
420 N. Wabash
Chicago, IL 60614
Email: smeisels@erikson.edu

Samuel J. Meisels is president of the Erikson Insititute. An emeritus professor of education at the University of Michigan, Dr. Meisels's research concerns the development of alternative assessments from birth through grade 5, policy studies of early intervention, and research regarding the developmental consequences of high risk birth. He has co-authored *The Handbook of Early Intervention*, the "Work Sampling System," and the "Early Screening Inventory-Revised," among many other publications.

Sally Atkins-Burnett is an assistant professor at the University of Toledo. She earned her Ph.D. from the University of Michigan. She has been a major contributor to the NCES-sponsored Early Childhood Longitudinal Study and has also worked on the development of instruments used in the Study of Instructional Improvement, a national longitudinal study of the impact of school improvement efforts on instruction and student performance. Her research interests include assessment in early childhood, social development, literacy instruction, and children with special needs.

Yange Xue is a Research Scientist at the National Center for Children and Families, Teachers College, Columbia University, New York, NY. She holds a doctorate in early childhood education from the University of Michigan. Her areas of specialization include early development and education and quantitative methodology.

Julie Nicholson received her PhD in Early Childhood Education from the University of Michigan. Her research interests include emergent literacy, the use of technology in early childhood classrooms, and the positive and negative consequences associated with different methods of early childhood assessment.

Donna DiPrima Bickel, is a Resident Fellow at the University of Pittsburgh's, Institute For Learning (IFL). Since 1999, Dr. Bickel has coordinated the IFL program in Content-Focused Coaching in elementary literacy. She has co-developed the professional development video and print materials used in preparing school district coaches to work with teachers using this model of professional development.

Seung-Hee Son is a doctoral student in Early Childhood Education at the University of Michigan. Her research interests include parenting and preschool influences on early language and literacy development, and policy issues in child development.

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler

University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 1897 times since March 14, 2003

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Education Policy Analysis Archives

Volume 11 Number 10 **March 14, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article if EPAA is credited and copies are not sold. EPAA is a project of the Education Policy Studies Laboratory.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Exploring the Achievement Gap Between White and Minority Students in Texas: A Comparison of the 1996 and 2000 NAEP and TAAS Eighth Grade Mathematics Test Results

Thomas H. Linton
Texas A&M University-Corpus Christi

Donald Kester
Texas A&M University-Corpus Christi

Citation: Linton, T. H. & Kester, D. (2003, March 14). Exploring the achievement gap between white and minority students in Texas: A comparison of the 1996 and 2000 NAEP and TAAS eighth grade mathematics test results, Education Policy Analysis Archives, 11(10). Retrieved [date] from http://epaa.asu.edu/epaa/v11n10/.

Abstract

The Texas Assessment of Academic Skills (TAAS) has been used to document and track an achievement gap between white and minority students in Texas. Some educators have credited the TAAS with fueling a drive to close the achievement gap while others suggest that TAAS scores may be misleading because of factors such as score inflation and a possible ceiling effect. The purpose of this study was to analyze the gap in mathematics achievement for eighth grade students. The study compared TAAS and National Assessment of Educational Progress (NAEP) test results to determine if the achievement gap between white, Hispanic, and African-American Students had narrowed between 1996 and 2000. Results indicate that TAAS mean scores increased significantly for all three ethnic groups between 1996 and 2000. Comparison of the TAAS test score frequency distributions for each ethnic group indicated that white students' scores shifted from the middle to the upper portion of the test score range while minority students' scores shifted from the lower to the middle and higher score range. Both white and minority students' TAAS test score distributions were significantly more negatively skewed in 2000 than in 1996. Comparisons between white and minority students' TAAS scores showed that white students had significantly higher scores than either Hispanic or African-American students in both 1996 and 2000. Comparison of mean score differences in 1996 and 2000 indicated that the achievement gap between white and minority students had narrowed. NAEP scores increased significantly from 1996 to 2000 for Hispanic students, but not for white or African-American students. However, test score distribution patterns showed small positive changes for all three ethnic groups. Comparisons between ethnic groups indicated that there were significant differences between white and minority students' scores in both 1996 and 2000. Comparison of mean score differences in 1996 and 2000 indicated that the achievement gap between Hispanic white students had narrowed slightly but that there was no change in the achievement gap between white and African-American students. Analysis of the TAAS test score distribution patterns indicated the likelihood that a ceiling effect had impacted students' scores. The evidence for a ceiling effect was strongest for white students. In 2000, 60.4% of white students had a TAAS score that fell in the top 10% of the score range. In contrast, there was no evidence of a ceiling effect for the NAEP. Mean score gains on the TAAS are only

partially substantiated by the NAEP data. Furthermore, there is a very strong possibility that a ceiling effect artificially restricted the 2000 TAAS scores for white students and created the illusion that the achievement gap between minority and white students had been narrowed.

The Texas Assessment of Academic Skills (TAAS) has been administered as a measure of student achievement in reading, mathematics, and writing in Texas since 1990. The tests have been praised for providing disaggregated data on different ethnic groups and requiring each group to meet the same standard of proficiency. The disaggregated scores have been used to document and track an achievement gap between the scores of white and minority students. Test results have typically been reported as percent of students passing (i.e. meeting minimum expectations) the TAAS (Texas Education Agency, 2000 b). These results show that passing rates for minority students have increased at a faster rate than the passing rates for white students.

Table 1
TAAS Grade 8 Mathematics Test: A Comparison of Percent Passing for 1996 and 2000 by Ethnic Group

Ethnic Group	Percent Passing		Percent Gain
	1996	2000	
African American	43.8	80.9	37.1
Hispanic	51.4	85.5	34.1
White	77.7	94.8	17.1

Many educators in Texas have credited the tests as fueling the drive to close this achievement gap. Several studies have cited the increased percent of minority students passing the TAAS as evidence that the achievement gap is being closed (Hurley, Chamberlain, Slavin, & Madden, 2001; Jerald, 2000; Jerald, 2001; Texas Education Agency, 2000 a). Other researchers disagree and suggest that increases in passing rates for minority students may be linked to other factors. Haney (2000) argues that increases in the TAAS passing rate for minority high school students is at least partially explained by higher dropout rates among minority students. Toenjes and Dworkin (2002) cite evidence that challenges Haney's assertions and conclude that dropout rates and special education exemptions do not explain the large increases in the TAAS passing rate.

It should be noted that none of these studies directly address the actual achievement level of students. Comparing the passing rate for minority and white students does not provide information about comparative achievement levels. Rather, the passing rate is simply the percentage of students that have attained the minimum achievement level deemed necessary to "pass" the test.

Rather than depending solely on TAAS data, some researchers have used National Assessment of Educational Progress (NAEP) results to analyze the achievement levels of Texas students. These studies have consistently found either no gain or small achievement gains for both minority and white students, but have been unable to substantiate the large gains reported on the TAAS. Amrein and Berliner, after examining test results for 18 states with high-stakes testing programs, reported that student achievement remained at the same level or went down after the high-stakes testing policies were instituted. Camilli, using a cohort analysis of gains on the NAEP math test from grade 4 to grade 8 found that Texas ranked 17th among 35 states. In a study that directly compared TAAS and NAEP test results (Klein, Hamilton, McCaffrey, & Stecher, 2000), TAAS data indicated that the achievement gap between minority and white students was closing while NAEP data did not. Klein, et. al. suggested that the large gains of minority students relative to white students on the TAAS were misleading and could be due, at least in part, to a ceiling effect and teaching to the test. However, Klein's study compared 1992 and 1996 NAEP scores with TAAS data for 1994 and 1998. Because of the disparity in the test administration dates for the NAEP and the TAAS, questions were raised about the conclusions of the study. Furthermore, Klein's study analyzed mean achievement gains for the TAAS and the NAEP but did not examine the actual distributions of test scores for evidence of a ceiling effect.

The National Center for Educational Statistics, in its report, *The Nation's Report Card: Mathematics 2000,* (U. S. Department of Education, 2001a) released national and state reports of NAEP scores for eighth grade mathematics in August 2001. To date, there have been no studies comparing the NAEP 2000 results and the TAAS results for Texas.

Purpose of the Study

The purpose of the present study is to present an analysis of the Texas TAAS and NAEP results for 1996 and 2000 and to explore the mathematics achievement gap between eighth grade minority students and white students in Texas. Specific research questions for the study are:

- Did Texas eighth grade African American, Hispanic, and white student math scores on the TAAS and NAEP increase significantly between 1996 and 2000?
- DO TAAS and NAEP data show that the achievement gap between white and minority students decreased from 1996 to 2000?

- Is there evidence that a ceiling effect artificially restricted the distribution of students' scores on the TAAS or the NAEP?

Methods

A causal-comparative research design was used to analyze the variables in the study. The sample for the TAAS data consisted of all African American, Hispanic, and white students in Texas who were in the Texas Education Agency's accountability subset of TAAS scores in grade 8 in 1996 and 2000 (TEA, 2000 b). (Note 1) The sample for the NAEP-Texas data consisted of a random sample of approximately 2,500 eighth grade students selected according to NAEP specifications in 1996 and 9,600 were selected in 2000. The national results for the NAEP test were included for comparative purposes. (Note 2)

Two sets of analyses were used to compare scores within and between ethnic groups. First, confidence intervals were calculated and used to analyze differences between mean scores. A *d* statistic (Green & Akey, 2000) was calculated to measure effect size for differences between mean scores. (Note 3) Second, a chi-square goodness-of-fit test was used to compare TAAS 1996 and 2000 test score distributions and NAEP 1996 and 2000 test score distributions for white and minority students. The Cramer's V coefficient (Green & Akey, 2000) was calculated to determine effect size for the chi-square analysis. (Note 4)

Research Question #1

Did Texas eighth grade African American, Hispanic, and white student math scores on the TAAS and NAEP increase significantly between 1996 and 2000?

The 1996 and 2000 TAAS scores and NAEP scores for each ethnic group were analyzed in two ways. First, mean scores for each ethnic group were compared to determine if there were significant differences between the 1996 results and the 2000 results. Second, the test score distributions for each ethnic group for 1996 and 2000 were compared to determine changes in the score distribution pattern over time.

TAAS Results

The comparison of 1996 and 2000 mean scores for each ethnic group on the TAAS is presented in Table 2. Results indicate that:

- Mean scores increased significantly for each ethnic group from 1996 to 2000.
- Effect sizes as measured by the d-statistic were moderate, ranging from .472 (white) to .646 (African American).
- Mean score gain for white students less than two-thirds as large as the gain for Hispanic and African American students.

Table 2
Comparison of TAAS TLI Mean Scores for 1996 and 2000 By Ethnic Group

Group	Ethnicity	Mean TLI Scores		Gain	Effect Size (d Statistic)
		1996	2000		
TAAS	African American	65.0	77.2	+12.2*	.916
	Hispanic	67.8	79.3	+11.5*	.894
	White	77.2	84.2	+7.0*	.667

* $p < .05$

The second set of TAAS analyses compared the 1996 and 2000 score distributions for each ethnic group. A chi-square analysis indicated that the score distributions for each ethnic group changed significantly from 1996 to 2000. Results of the analyses are presented in Tables 3 and 4.

- For each ethnic group, the distribution of scores in 1996 was significantly different from the distribution of scores in 2000.
- Effect size as measured by the phi coefficient indicated that there was a large gain for African American and a moderate gain for Hispanic students than for white students
- Scores for Hispanic and African American students tended to increase from the lowest score range to middle and upper portion of the score range.
- Scores for white students were concentrated in the upper portion of the score range.

Table 3
TAAS TLI Score Distributions in Percent for 1996 and 2000 by Ethnic Group

Ethnic Group		TLI Score					
		0-69	70-74	75-79	80-84	85-89	90-94
African Am.	1996	56.2	11.8	12.0	10.6	7.4	2.0
	2000	19.1	11.3	18.4	24.4	20.9	6.0
Hispanic	1996	48.6	11.9	13.1	12.9	10.3	3.3
	2000	14.5	9.0	16.3	24.9	26.0	9.4
White	1996	22.3	10.0	13.9	19.6	23.2	11.0
	2000	5.2	3.9	9.1	21.4	37.9	22.5

Table 4
Comparison of TAAS Score Distributions for 1996 and 2000
Within Each Ethnic Group

Test	Group	Significance Level	Effect Size (Cramer's V)
TAAS	White 1996-2000	<.01	.332
	African American 1996-2000	<.01	.405
	Hispanic 1996-2000	<.01	.410

NAEP Results

Two sets of NAEP mean scores were reported: NAEP results for Texas and NAEP results for the nation. The comparison of 1996 and 2000 mean scores within each ethnic group is presented in Table 5.

Results indicated that:

- For the NAEP-Texas, mean scores increased significantly for Hispanic students but not for African American or white students.
- For the National NAEP, mean scores increased significantly for white students but not for Hispanic or African American students.

Table 5
Comparison of NAEP Mean Scores for 1996 and 2000 within each Ethnic Group

Group	Ethnicity	Mean Scale Scores		Gain
		1996	2000	
NAEP-Texas	African American	249	252	+3.0
	Hispanic	256	266	+10.0*
	White	285	288	+3.0
NAEP-Nation	African American	242	246	+4.0
	Hispanic	250	252	+2.0
	White	281	285	+4.0*

* $p < .05$

The second set of analyses used a chi-square analysis to compare the 1996 and 2000 score distributions for each ethnic group. Results of the analyses are presented in Tables 6 and 7.

- For the NAEP-Texas, there was a significant change in the score distributions of all three ethnic groups from 1996 to 2000.
- The effect size (phi coefficient) for the NAEP-Texas was very small for white (.067) and African American students (.127) and slightly larger for Hispanic students (.176). This indicates that there was a more of a change in the Hispanic students' score distribution than for white or African American students.
- For the National NAEP, white students' score distribution changed significantly from 1996 to 2000 but the score distributions for Hispanic and African American students did not.
- The effect size for the National NAEP was very small for all three groups (.030 to .053), indicating that the

changes from 1996 to 2000 were minimal.

Table 6
Distribution of NAEP Score in Percent for 1996 and 2000 by Ethnic Group

Ethnicity	Year	Below Basic	Basic	Proficient	Advanced
Texas					
African Am.	1996	69	26	4	1
	2000	60	34	6	0
Hispanic	1996	58	34	7	1
	2000	41	45	13	1
White	1996	22	45	29	4
	2000	17	46	33	4
Nation					
African Am.	1996	73	23	4	0
	2000	68	27	5	0
Hispanic	1996	63	29	7	1
	2000	60	31	8	1
White	1996	27	43	25	5
	2000	23	43	28	6

Table 7
Comparison of NAEP Score Distributions for 1996 and 2000 within Ethnic Group

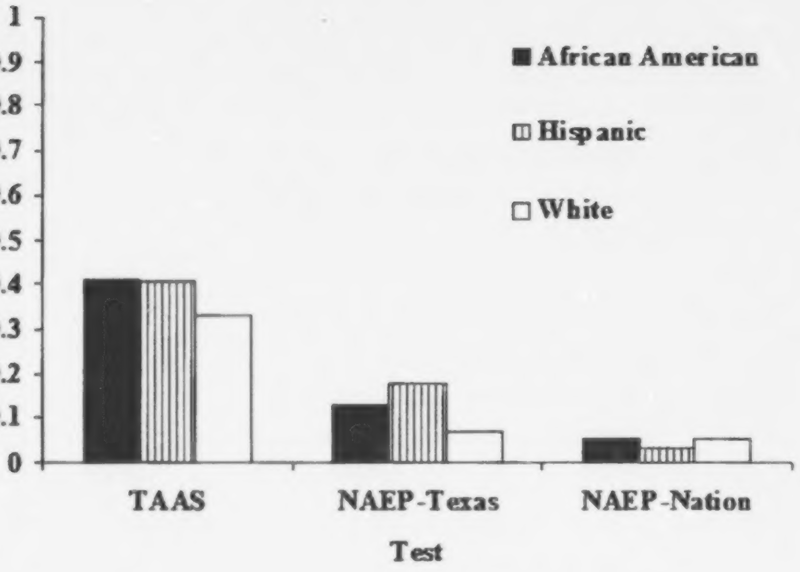
Test	Group	Significance Level	Effect Size (Cramer's V)
NAEP-Texas	White 1996-2000	<.05	.067
	Hispanic 1996-2000	<.01	.176
	African Am. 1996-2000	<.05	.127
NAEP-National	White 1996-2000	<.01	.052
	Hispanic 1996-2000	N.S.	.030
	African Am. 1996-2000	N.S.	.053

Conclusions

Comparison of the TAAS results and NAEP-Texas results show a significant mean score increase with a moderate to large effect size for all three ethnic groups on the TAAS. In contrast, only Hispanic students had a significant mean score increase on the NAEP-Texas, with a small effect size. The Hispanic students' increase on the Texas NAEP was not reflected on the National NAEP, indicating that the NAEP-Texas result was not part of a national trend.

The score distributions for both the TAAS and the NAEP-Texas changed significantly from 1996 to 2000. Effect sizes for changes in the TAAS score distributions were much larger than those found for the NAEP-Texas (Figure 1). In addition, the pattern of change in the distributions was different for the TAAS than for the NAEP-Texas. The TAAS distributions for African American and Hispanic students showed a very large decrease in students who failed the test (i.e. scored below 70 TLI) and an increase in scores in the middle to high range. For white students, the percentage at the lower and middle range decreased and the percentage at the top of the test score range showed a very large increase. In contrast, the NAEP-Texas distributions had changes primarily at the lower end of the score range for all three ethnic groups; that is, from "below basic" to "basic".

Figure 1.
Effect Size Comparison for 1996 and 2000 Score Distributions for TAAS and NAEP



Research Question 2

Do TAAS and NAEP data show that the achievement gap between white and minority students decreased from 1996 to 2000?

The 1996 and 2000 TAAS scores and NAEP scores for white and minority students were compared in two ways. First, mean scores for white students were compared with mean scores for African American and Hispanic students to determine if there were significant differences. Second, the distribution of test scores for white students and minority students were compared to determine if the distribution patterns became more similar over time.

TAAS—Comparison of White and Minority Scores

The comparison of mean scores for white and minority students on the TAAS presented in Table 8. Confidence intervals were used to determine the statistical significance of differences between white and minority students' mean scores in 1996 and in 2000.

- Mean scores for white students were significantly higher than African American students in both 1996 and 2000.
- The difference in mean scores for white and African American students was larger in 1996 than in 2000.
- Mean scores for white students were significantly higher than Hispanic students in both 1996 and 2000.
- The difference in mean scores for white and African American students was larger in 1996 than in 2000.
- Effect sizes for white vs. African American students was large while effect size for white vs. Hispanic students was moderate.

Table 8
Comparison of TAAS Mean TLI Score For White and Minority Students

Comparison Group	Mean Difference	Effect Size (d Statistic)
African American vs. White 1996	12.2*	.917
African American vs. White 2000	7.0 *	.824
Hispanic vs. White 1996	9.4 *	.682
Hispanic vs. White 2000	4.9 *	.553

* p <.01

A chi-square analysis compared the score distribution for white students with the score distributions for Hispanic students and African American students. Results of the analyses are presented in Table 9.

- The score distribution for white students was significantly different from the score distributions for African American students both 1996 and 2000.
- The score distribution for white students was significantly different from the score distributions for Hispanic students both 1996 and 2000.
- Effect sizes changed very little from 1996 to 2000. This indicated that, although the distribution patterns changed, the differences between white students' scores and minority students' scores were relatively unchanged from 1996 to 2000.

Table 9
Comparison of TAAS Score Distributions For White and Minority Students

Comparison Group	Significance Level	Effect Size (Cramer's V)
African American vs. White 1996	<.01	.331
African American vs. White 2000	<.01	.333
Hispanic vs. White 1996	<.01	.313
Hispanic vs. White 2000	<.01	.286

NAEP—Comparison of White and Minority Scores

The comparison of mean scores for white and minority students on the NAEP-Texas and the National NAEP are presented in Table 10. Confidence intervals were reported to show the statistical significance of differences between white and minority students' mean scores in 1996 and 2000. For the NAEP-Texas:

- White students' mean scores were significantly higher than African American and Hispanic students' mean scores both in 1996 and in 2000.
- The differences between white and African American students' mean scores remained large and relatively unchanged from 1996 to 2000.
- The difference between white and Hispanic students' mean scores decreased from 1996 to 2000.

For the National NAEP, white students' mean scores were significantly higher than African American and Hispanic students' mean scores both in 1996 and in 2000. The differences were large and did not change appreciably from 1996 to 2000.

Table 10
Comparison of NAEP Mean Scale Scores For White and Minority Students

Test	Comparison Group	Mean Difference
NAEP-Texas	African American vs. White 1996	36*
	African American vs. White 2000	36 *
	Hispanic vs. White 1996	29 *
	Hispanic vs. White 2000	22 *
NAEP-National	African American vs. White 1996	39 *
	African American vs. White 2000	39 *
	Hispanic vs. White 1996	31 *
	Hispanic vs. White 2000	31 *

* p<.01

A chi-square analysis compared the score distribution for white students with the score distributions for Hispanic students and African American students. Results of the analyses are presented in Table 11.

- On the NAEP-Texas, the score distributions for Hispanic and African American students were significantly different from that of white students in both 1996 and 2000.
- On the NAEP-Texas, the effect size for the comparison of Hispanic students' and white students' scores in 2000 was smaller than in 1996. In contrast, the effect size for the comparison of African American students' and white students' scores were unchanged from 1996 to 2000.
- On the National NAEP, the score distributions for African American and Hispanic students were significantly different from the score distribution for white students in both 1996 and 2000.
- On the National NAEP, effect size was moderate and there was little change from 1996 to 2000.

Table 11
Comparison of NAEP Score Distributions For White and Minority Students

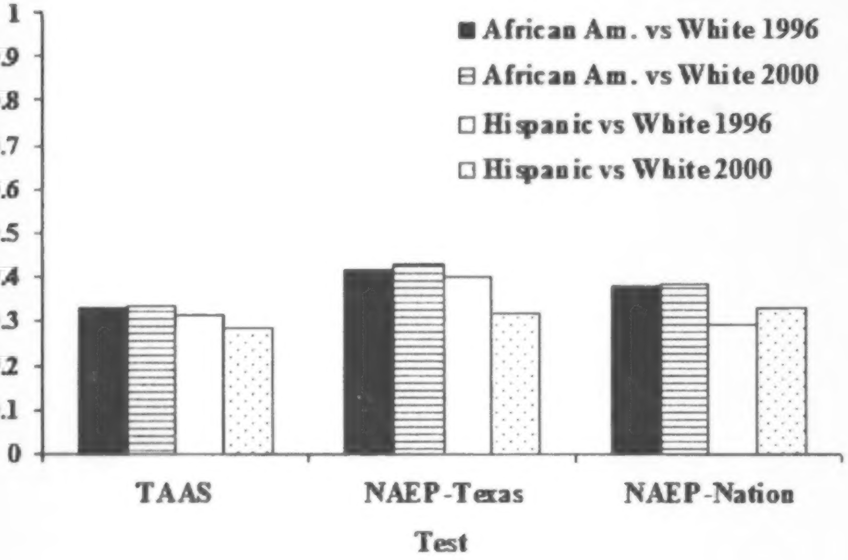
Test	Comparison Group	Significance Level	Effect Size (d Statistic)
NAEP-Texas	African Am. vs. White 1996	<.01	.415
	African Am. vs. White 2000	<.01	.428
	Hispanic vs. White 1996	<.01	.402
	Hispanic vs. White 2000	<.01	.318
NAEP-National	African Am. vs. White 1996	<.01	.379
	African Am. vs. White 2000	<.01	.383
	Hispanic vs. White 1996	<.01	.291
	Hispanic vs. White 2000	<.01	.327

Conclusions

The TAAS results show that the difference in mean scores for white and minority students was smaller in 2000 than in 1996. This would seem to indicate that minority students were closing the achievement gap in eighth grade mathematics. However, results of the NAEP-Texas offer only partial support for this conclusion. On the NAEP-Texas, the difference for white and African American students was unchanged from 1996 to 2000. The difference in NAEP-Texas mean scores for Hispanic and white students was smaller in 2000 than in 1996, although the difference was still large. In contrast, results from the National NAEP indicated that the difference between Hispanic and white students' mean scores actually increased slightly.

Comparison of the score distributions of white and minority students presents a similar result. Comparison of the score distributions of white and African American students yielded similar effect sized in 1996 and 2000. In contrast, the comparison of white and Hispanic students showed that the effect size decreased from 1996 to 2000 (Figure 2).

Figure 2.
Comparison of Effect Size for White vs. Minority Score Distributions



The finding that the effect size for comparisons of minority and white students is larger on the NAEP-Texas than on the TAAS show that the disparity between minority and white students is greater on the NAEP-Texas. That is, the achievement gap is more evident on the NAEP-Texas than on the TAAS both in 1996 and in 2000. However, on both tests, the achievement gap between Hispanic and white students was smaller in 2000 than in 1996. The fact that Hispanic students do not show similar gains nationally indicates that this is not part of a national trend. The disparity between the NAEP-Texas and National NAEP may be an indication that Hispanic students in Texas are beginning to close the achievement gap in eighth grade mathematics.

Research Question 3

Is there evidence that a ceiling effect artificially restricted the distribution of students' scores on the TAAS or the NAEP?

TAAS Scores

The analysis of TAAS mean score gains for each ethnic group show that white students gained only 7.0 TLI points from 1996 to 2000 while African Americans gained and Hispanic students gained 11.5. Since the largest percentage of white students scores were in the upper 10% of the score range in both in 1996 and 2000, the gains for their highest scoring students were limited to the maximum score possible on the test. The likely result is that they were not able to show their true achievement level because the maximum score (ceiling) of the test artificially limited their scores. If this were the case, comparison of their scores with those of minority students (whose opportunity for gain was not as restricted) would create the appearance that the lower scoring students were achieving at a greater rate and therefore closing the achievement gap.

A second analysis looked at the distribution of scores for each ethnic group for evidence of a ceiling effect. This analysis, presented in Table 12, gives the percent of students in each ethnic group with TLI scores in the upper 10% of the score range (i.e. a TLI of 85 to 94). The table shows that in 1996, 34% of the white students had test scores in the upper 10% of the TAAS score range. This increased to 60% in 2000. The fact that the white students have the largest percentage of students in the upper range indicates that the score range for these students is more restricted by the maximum test score (the test ceiling).

The dramatic difference in the score distributions for white and minority students provides support for the hypothesis that a ceiling effect has restricted white student' scores to a greater degree than it has restricted Hispanic and African American students scores. If this hypothesis is correct, the result would be an artificial narrowing of the achievement gap between white and minority students' eighth grade math test scores.

Table 12
Percent of Students with TAAS TLI Scores Greater than 84

Math	TLI Scores		Cumulative Total
	85-89	90-94	
White 1996	23.2	11.0	34.2
White 2000	37.9	22.5	60.4
African American 1996	7.4	2.0	9.4
African American 2000	20.9	6.0	26.9
Hispanic 1996	10.3	3.3	13.6
Hispanic 2000	26.0	9.4	35.4

NAEP Scores

NAEP-Texas and National NAEP results showed that the mean scores for all three ethnic groups were near the middle of the test range (0 to 500). An analysis of the score distributions for the NAEP-Texas and the National NAEP (Table 13) show that student gains have been primarily at the lower range of the test (from "below basic" to "basic"), with little change in the percent of students achieving the "advanced" range. There is no evidence to support the hypothesis that there is a ceiling effect for either the NAEP-Texas or the National NAEP results.

Table 13
Percent of Students with NAEP Scores Above Basic

Ethnicity	Year	Proficient	Advanced	Cumulative Total
NAEP-Texas				
African American	1996	4%	1%	5%
	2000	6%	0%	6%
Hispanic	1996	7%	1%	8%

	2000	13%	1%	14%
White	1996	29%	4%	33%
	2000	33%	4%	37%
NAEP-National				
African American	1996	4%	0%	4%
	2000	5%	0%	5%
Hispanic	1996	7%	1%	8%
	2000	8%	1%	9%
White	1996	25%	5%	30%
	2000	23%	43%	34%

Discussion

White, African American, and Hispanic students all had large and statistically significant gains on the TAAS from 1996 to 2000. Comparison of white and minority students' scores show that white students had significantly higher TAAS scores than African Americans and Hispanics in 1996 and 2000, but that the differences were smaller in 2000. These results were consistent both for analysis of mean scores and analysis of the test distributions for each student group.

NAEP results were not consistent with the TAAS results. Hispanic students had a mean score gain from 1996 to 2000, but white and African American students did not. When white students' NAEP scores were compared to minority students' scores, the difference between Hispanic and white students' scores decreased from 1996 to 2000 but the difference between African American and white students' scores did not. These results were consistent for analysis of mean scores and test distributions for each student group.

In summary, the large student gains on the TAAS, which is a minimum skills test tailored specifically to the Texas mathematics curriculum, are only partially substantiated by the smaller gains on the NAEP, which is a more general and more difficult test of mathematics. While an explanation of the reasons for the differences in the TAAS and NAEP results is beyond the scope of this research, the authors' experience in Texas public schools suggests two likely answers. First, teaching to the TAAS is widespread and pervasive in Texas schools. Release versions of the TAAS are available from the Texas Education Agency (along with scoring services that mimic actual TAAS reports) as are a variety of commercially developed practice and test preparation materials. It is common practice for schools to administer one or more "practice TAAS" tests in the fall and use the results to guide instruction in preparation for the state-mandated TAAS testing in the spring. Second, Texas teachers and principals are evaluated, in part, on their students' success (or lack of success) on the TAAS. These factors create very strong pressure to teach to the test. It is likely that score inflation is a significant factor in the large gains that have been consistently reported for the TAAS.

TAAS data for 2000 revealed that differences between white and minority students' scores had decreased when compared to 1996. Other studies have considered this as evidence that the achievement gap between white and minority students is being narrowed. However, analysis of the distribution of scores for each ethnic group reveal that over 60% of white students scored in the upper 10% of the test score range while about 27% of African American students and 35% of Hispanic students scored in this range.

Since a larger percentage of white students than minority students achieved the maximum score on the test (the test ceiling), white students' scores likely underestimated their true achievement level. That is, the ceiling effect has artificially restricted white students scores and created the illusion that the achievement gap has been narrowed. The presence of a ceiling effect casts doubt on the validity of claims that the achievement gap between white and minority students has been narrowed. A more reasonable interpretation of the available data is that because the test ceiling has differentially affected the scores for white, Hispanic, and African American students, TAAS results cannot be used to determine whether or not the achievement gap has been narrowed.

Analysis of the NAEP-Texas score distribution suggests that the achievement gap for African American and white students has not changed between 1996 and 2000. However, the gap between white and Hispanic students' scores did narrow, although the change was small when compared to the TAAS. Comparison of NAEP-Texas and National results indicates that this change was a Texas phenomena and was not found in the National NAEP data. This finding does indicate that Texas has been partially successful in narrowing the achievement gap between Hispanic and white students.

The results of this study have implications beyond the TAAS and the State of Texas. The national emphasis on high standards and the use of high stakes criterion-referenced tests to measure progress toward those standards have become commonplace in public education. Many states depend solely on high stakes test results for making far-reaching decisions about the content and formulation of curricula, the funding of educational initiatives, and the development of educational policy. Any state that uses a high stakes test to measure progress toward state standards must be aware of the twin dangers of test score inflation and ceiling effects. Both can lead to invalid interpretation of test scores and erroneous conclusions about student achievement. The use of comparative data such as the NAEP are vital to ensure that the test data used by state and national decision-makers presents an accurate picture of the educational achievement of their students.

Notes

1. TAAS data were ordered as a customized report of frequency distributions by ethnic group. The data set for each ethnic group consisted of a frequency count of the number of students by Texas Learning Index (TLI) score together with the mean, standard deviation, and SEM of the distribution. The TLI score is a scaled score derived specifically for the TAAS and is not comparable to other scaled scores. A complete description of the derivation of the TLI is contained in the TAAS Technical Digest available from the following Texas Education Agency web site (TEA, 2000 c).
2. All data for the NAEP-Texas and National NAEP were obtained from the following NCES reports: *The nation's report card: Mathematics 2000.* (U.S. Department of Education, 2001a) and *The nation's report card: state mathematics 2000, report for Texas* (U.S. Department of Education, 2001b)
3. The *d* statistic is the ratio of the mean difference between two groups divided by the pooled standard deviation. A value of .2, .5, or .8 is generally interpreted as small, medium, or large effect size, respectively.
4. The Cramer's V coefficient is a rescale of the phi coefficient and has a range between 0 and 1. A Cramer's V of .1, .3, or .5 is generally interpreted as small, medium, or large effect size, respectively.

References

Amrein, A. L. and Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10(18).* Retrieved March 14, 2003, from http://epaa.asu.edu/epaa/v10n18/.

Camilli, G. (2000). Texas gains on NAEP: points of light? *Education Policy Analysis Archives, 8(42).* Retrieved March 14, 2003, from http://epaa.asu.edu/epaa/v8n42/.

Green, S. & Akey, T. (2000). *Using SPSS for windows: Analyzing and understanding data* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Haney, W. (2000, August). The myth of the Texas miracle in education. *Education Policy Analysis Archives, 8*(41). Retrieved March 14, 2003, from http://epaa.asu.edu/epaa/v8n41/.

Hurley, E., Chamberlain, A., Slavin, R., and Madden, N. (2001). Effects of success for all on TAAS reading scores: A Texas statewide evaluation. *Phi Delta Kappa, 82*(10), 750-756.

Jerald, C. (2000, January 13). The state of the states. *Education Week, 19*(18), 62-65.

Jerald, C. (2001). Real results, remaining challenges: The story of Texas education reform. *The Education Trust.* Retrieved March 14, 2003, from http://www.brtable.org/document.cfm/532.

Klein, S., Hamilton, L., McCaffrey, B., & Stecher, B. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives 8(49).* Retrieved March 14, 2003, from http://epaa.asu.edu/epaa/v8n49/.

National Center for Educational Statistics (2001a). *The nation's report card: Mathematics 2000* (NCES Document No. 2001-517). Washington, D.C: Office of Educational Research and Improvement, U. S. Department of Education.

National Center for Educational Statistics (2001b). *The nation's report card: state mathematics 2000, report for Texas* (NCES Document No. 2001-519 TX). Washington, DC: Office of Educational Research and Improvement, U. S. Department of Education.

Texas Education Agency (2000 a, April 20). *State's exit-level TAAS performance sets another record: 10th graders hit 90 percent mark on two sets of tests.* Austin, TX: Author.

Texas Education Agency, Office of Policy Planning and Research. (2000 b, April). *The 2000 accountability rating system for Texas public schools and school districts.* Austin, Texas: Author.

Texas Education Agency (2000 c). *TAAS Technical Digest.* Retrieved January 2, 2002, from http://www.tea.state.tx.us/student.assessment/researchers.html

Toenjes, L. and Dworkin, A. G. (2002). *Are increasing test scores in Texas really a myth, or is Haney's myth a myth?* Education Policy Analysis Archives 10(17). Retrieved March 14, 2003, from http://epaa.asu.edu/epaa/v10n17/.

About the Authors

Dr. Tom Linton is an associate professor in the Educational Leadership Doctoral Program at Texas A&M University-Corpus Christi. His experience includes 25 years in the Corpus Christi ISD Office of Research Testing, and Evaluation and with various other school districts as an outside consultant and as a project evaluator. He has also served on several advisory committees for the Student Assessment Division of the Texas Education Agency. His e-mail address is: tlinton@falcon.tamucc.edu.

Dr. Don Kester is an associate professor in the Educational Leadership Doctoral Program at Texas A&M University-Corpus Christi. His experience includes 27 years as an administrative consultant in the areas of educational program management, evaluation, research, and student assessment with the Los Angeles County Office of Education. His e-mail address is: don.kester@mail.tamucc.edu.

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)

J. Félix Angulo Rasco (Spain)

Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Este artículo ha sido consultado 1784 veces desde el 21 de marzo 2003

archives | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 11 Number 11 **marzo 21, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 2003, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Análisis del concepto de "la docencia" en profesores universitarios: Un estudio cualitativo

Gustavo Hawes Barrios
Sebastián Donoso Díaz
Universidad de Talca, Chile

Citation: Hawes Barrios, G. & Donoso Díaz, Sebastián. (2003, marzo 21). Análisis del concepto de "la docencia" en profesores universitarios: Un estudio cualitativo, *Education Policy Analysis Archives*, *11*(11). Retrieved [date] from http://epaa.asu.edu/epaa/v11n11/.

Resumen
El trabajo explora las concepciones de aprendizaje y enseñanza en docentes universitarios sin formación pedagógica. Como marco de referencia utiliza el modelo de las teorías implícitas, y se apoya en teorías psicológicas del aprendizaje para dar consistencia al proceso analítico. Las observaciones fueron hechas entre docentes de una universidad pública de Chile. Los resultados muestran una visión más bien mecanicista del aprendizaje, asociada a una concepción de la enseñanza apoyada en el concepto de la transmisión, si bien aparecen algunos intentos de ligar la temática con el procesamiento de datos. La visión del profesor profesional es más romántica que profesional.

Abstract
This article explores the conceptions of learning and teaching among university teachers with no pedagogical qualifications. The model of implicit theories is used as a frame of reference, and the analytical process is supported by psychological theories of learning. Observations were carried out in a public university in southern Chile. Results evidence a rather mechanistic vision of learning, associated with a vision of teaching resting on the concept of transmission, although there are some efforts to link this conception with data processing. The vision of the professional teacher is more romantic than professional.

1. Presentación.

El trabajo explora los conceptos de la docencia en docentes universitarios que cumplen esa función. Es sabido que la mayoría de los docentes en las universidades no posee formación sistemática en el campo didáctico, sino que se han incorporado a las instituciones a partir de sus calificaciones académicas y antecedentes profesionales o investigativos.

Diversas consideraciones hacen que explorar estos conceptos sea una tarea relevante. Según Smith (1999) la conferencia –modelo de la clase- fue diseñada inicialmente como una forma de compensar la escasez de textos; sin embargo, logró sobrevivir a su propósito. MacFarlane (1999) plantea que existen tres áreas en las cuales es urgente la investigación: estructurar el conocimiento para su acceso interactivo, desarrollar tecnologías de apoyo al aprendizaje y, apoyar las comprensiones tácitas. Como se verá, sacar a la luz las comprensiones tácitas es equivalente a conocer las teorías implícitas y hacer explícito lo implícito.

Fernández (1995) plantea la necesidad de dilucidar la relación sistémica entre las prácticas pedagógicas, la descripción y análisis de esas prácticas, y los supuestos teóricos sobre los que se sustentan. Las concepciones que tienen los enseñantes acerca de los procesos de enseñar, de aprender y las funciones asociadas, son decisivas para comprender las prácticas efectivas de docencia y también para iniciar procesos de cambio a partir de esos conocimientos y pre-concepciones, sobre las que se estructura un complejo marco de potenciales acciones.

También existen razones de orden práctico que avalan la importancia de un trabajo como el presente. Una se relaciona con la responsabilidad social de la universidad, en cuanto al tipo y calidad de la formación que entregan a los jóvenes y a los que acreditará posteriormente para el ejercicio de profesiones. Matricularse en un programa de formación constituye un contrato por el cual la institución se compromete a enseñar eficientemente los conocimientos y destrezas necesarios para un ejercicio profesional determinado.

Otro elemento que contribuye a sustentar esta propuesta es el correcto desempeño, es decir, con el ethos mismo de la academia. Es propio del maestro atender a que sus estudiantes aprendan; lo impropio es que el docente se desentienda del éxito o fracaso de los alumnos. En este sentido, la universidad es más que un mero agregado de individualidades como podría sugerir un concepto mercantilista de la organización que campea en nuestros tiempos, es una comunidad empeñada en la búsqueda de los valores superiores del espíritu (verdad, bien, belleza) que conlleva que los actores no tengan el derecho a desentenderse del logro o fracaso de los jóvenes estudiantes.

En una perspectiva de mercado es posible afirmar la relevancia del conocimiento que se busca producir con el propósito de hacer más eficientes y efectivas las prácticas docentes al interior de la universidad, contribuyendo a elevar la competitividad institucional, incrementando la atracción sobre los mejores estudiantes, reteniendo a una mayor cantidad de alumnos y así mejorando los ingresos económicos institucionales, ejerciendo una atracción también sobre los mejores docentes y constituyéndose en una alternativa atractiva tanto para estudiar como para trabajar.

El trabajo fue llevado a cabo durante el año académico 2002 en una universidad pública, en una ciudad de tamaño mediano en el sur de Chile. Se entrevistó a 15 académicos de las diferentes facultades e institutos universitarios, pertenecientes a las diversas jerarquías académicas, con el propósito de dar cuenta de la mayor cantidad de diferencias en las percepciones.

La universidad en la que se llevó a cabo el trabajo es de tamaño mediano, con unos 5000 estudiantes de pregrado distribuidos en más de una docena de programas de formación profesional; existen también programas de postgrado del nivel de magister y doctorado, estos últimos iniciados hace un par de años. Los estudiantes de pregrado provienen fundamentalmente del ámbito geográfico próximo, el que se caracteriza por estar entre las más deprimidas económica, cultural y educacionalmente del país. La universidad ha hecho grandes esfuerzos e inversiones por mejorar la calidad de la atención que se presta a los estudiantes. El énfasis en la docencia es una de las señales que se han entregado al respecto.

2. Antecedentes teoréticos.

Del punto de vista teorético, la investigación se enmarca en el programa de investigación (Nota 1) del pensamiento de los profesores. El enfoque de las teorías implícitas es un referente fundamental de la investigación, y los conceptos teoréticos en torno al aprendizaje y la enseñanza actuan como organizadores de la información que se levante.

2.1. Teorías implícitas .

Bajo este modelo se asume que toda acción sistemática llevada a cabo por los docentes se sostiene en un conjunto de concepciones que configuran una estructura denominable "teoría". El hecho de que no sean teorías conscientemente elaboradas o adoptadas las hace implícitas (Dienes y Perner, 1999). El término implícito se asocia a su contrario, explícito. Esta distinción es aplica a las representaciones cognoscitivas.

Se asume, de acuerdo a Dienes y Perner (1999), que el conocimiento es una actitud hacia una proposición que es verdadera, la cual por su parte predica una propiedad de alguna entidad. Respecto de ella podemos tener conocimiento de diferente nivel, tal que "*si se conoce un aspecto superior explícitamente, entonces cada nivel interior también debe ser conocido explícitamente*". De esta manera, el nivel más alto de conocimiento explícito marca el punto en que el restante "conocimiento" de nivel superior se transforma en implícito.

Vande (2001) propone que las teorías implícitas son conjuntos de "creencias" que sostienen las personas, y que éstas difieren considerablemente en las teorías implícitas o creencias que sostienen respecto de los atributos personales, de las destrezas interpersonales, la inteligencia, la personalidad, el carácter moral. La relevancia de este enfoque radica en que las acciones que lleven a cabo los docentes se basarán en creencias respecto de los atributos personales de los estudiantes, su inteligencia, personalidad, antecedentes, capacidad para aprender, incluyendo un cierto grado de "profecía" respecto de su éxito en la tarea de aprender: lo que se denomina el "efecto Pygmalión", que tiene que ver con las expectativas –no conscientes o no explícitas- de los profesores respecto de sus estudiantes (Arancibia, Herrera y Strasser, 2000).

2.2. Concepto de aprendizaje en el marco de la filosofía y la psicología.

El conocimiento y el aprendizaje con él, ha sido una preocupación constante de la indagación acerca de la realidad, no sólo en cuanto a saber qué conoce sino a qué puede conocer, es decir, cuáles son las condiciones bajo las cuales es concebible y aceptable hablar de conocimiento. El conocer es un hecho indubitable, no tendría sentido este texto si ello no fuera un dato de la causa.

Antes del surgimiento de la psicología como disciplina científica, en el marco del positivismo, este tema era abordado por los filósofos, siendo un capítulo esencial de todo sistema filosófico.

La aproximación en la filosofía.

En la tradición filosófica occidental moderna encontramos dos grandes líneas de pensamiento al respecto: una, correspondiente al modelo del criticismo kantiano de fines del siglo XVIII, y la otra que se identifica con la filosofía del positivismo de Comte en el siglo XIX.

El criticismo kantiano se propone superar la contraposición entre el idealismo de raigambre cartesiana (las ideas son innatas) y el empirismo inglés (no hay ideas, sólo sensaciones). Para ello, Kant propone en la Crítica de la Razón Pura las siguientes ideas fundamentales: (i) no conocemos directamente la realidad tal cual es (νούμενος) sino sólo en cuanto aparece o se manifiesta (φαι- νούμενος) al sujeto que conoce; (ii) el sujeto dispone de dos categorías denominadas "estética trascendental", que son el tiempo y el espacio, a partir de las cuales puede poner un orden en la multiplicidad de las impresiones que asaltan los sentidos; (iii) los datos ordenados por la conciencia estética son elaborados con la utilización de las "categorías trascendentales". Las doce categorías, que son formas a priori del entendimiento, son los "moldes" a través de los cuales se forman los conceptos empíricos, y sólo tienen validez aplicadas a las intuiciones suministrados por la sensibilidad. Estas categorías tienen que ver con la cantidad (unidad, pluralidad, totalidad), la cualidad (relación, negación, limitación), con la relación (sustancia y accidente; causa y efecto; acción recíproca), y con la modalidad (posibilidad-imposibilidad; existencia-no existencia; necesidad-contingencia).

El positivismo comtiano divide la historia del pensamiento de la humanidad en tres grandes etapas o estadios: el teológico, el metafísico, y el positivo. Mientras los dos primeros constituyen etapas primitivas de pensamiento, el estadio positivo es real, definitivo, constituye el punto más alto que puede alcanzar la inteligencia humana. En él no se hace referencia a dioses ni a sustancias incognoscibles al modo kantiano. La imaginación del pensamiento mágico y metafísico queda definitivamente subordinada a la observación empírica. La mente del hombre debe atenerse a las cosas observables: sólo se buscan los hechos y sus leyes. No es necesario buscar causas o principios de las sustancias, porque es un conocimiento inaccesible, precisamente porque no puede ser observado. El positivismo se atiene a lo que está "puesto", lo dado, el dato. La mente busca las leyes de los fenómenos mediante un proceso inductivo a partir de la observación.

Los dos elementos centrales de las posiciones señaladas: la actividad de la mente (criticismo kantiano) y lo observable como lo único real (positivismo) son las bases que sustentan la ciencia contemporánea y también la psicología del aprendizaje.

La aproximación en la psicología.

En el campo de la psicología del aprendizaje encontramos dos tradiciones principales: la del cognitivismo y la del conductismo, ambas con sus propias variantes internas.

El cognitivismo se asocia con Piaget, para quien el aprendizaje se explica como una estructura de significaciones y relaciones en estado de equilibración inestable. El proceso de equilibración es el que rige al aprendizaje: éste se produce solamente cuando existe un desequilibrio cognitivo entre las estructuras del sujeto y los estímulos provenientes del medio. El desequilibrio cognitivo se rige por los procesos de asimilación y acomodación. Asimilación implica la incorporación de nueva información a los esquemas mentales existentes previamente. Esta conlleva modificar los esquemas asimiladores en función de los elementos que habían sido asimilados, implicando que los conocimientos se pueden adaptar a las características reales de los sujetos. Ocasionalmente el estímulo provoca en el sujeto una disonancia cognitiva, un conflicto entre los esquemas preexistentes y la nueva información. Ante esto, hay dos respuestas: la no adaptativa que significa abandonar el esfuerzo por aprender (no cambian los esquemas); y la adaptativa, como toma de conciencia de la existencia de un conflicto, y el esfuerzo por resolverlo. Se detectan tres niveles de respuesta adaptativa: la "alfa", con toma de conciencia del conflicto pero sin cambio de esquemas; la "beta" en que el componente conflictivo se incorpora como un caso de variación, no implica reestructuración de los esquemas aunque reduce el grado de conflicto; la "gamma" en la cual hay una reestructuración del esquema donde la nueva información viene a integrarse.

El conductismo se plantea en sus orígenes (inicios del siglo 20) como una reacción frente a los métodos de estudio (introspección) y al objeto (conciencia). Watson proclama la necesidad de utilizar una metodología objetiva y el estudio de la conducta observable. Con ello hace negación de la existencia de fenómenos internos de conciencia, así como metodológicamente se sostiene que no existen métodos apropiados para estudiarlos. La idea central es la asociación estímulo-respuesta-contingencia de refuerzo. El aprendizaje es iniciado y controlado desde fuera del sujeto, de manera tal que la organización del medio (estímulos y contingencias de refuerzo) es la que produce el aprendizaje elicitando patrones de comportamiento observable. Se asume que las leyes del aprendizaje son completamente aplicables de manera homogénea y universal al conjunto de los organismos; igualmente, que cualquier estímulo puede asociarse con cualquier otro estímulo (equipotencialidad de los estímulos). El claro empirismo del enfoque lleva a concebir, como se indicó, la no existencia de los fenómenos de conciencia, ni ningún otro de carácter interno, negación que está fundada en la no-observabilidad de los mismos, y porque el modelo definitivamente no los necesita.

2.3. Una visión comparada.

La comparación de ambos modelos permitirá generar un patrón analítico para utilizarlo en el examen de las respuestas de los sujetos en la investigación que se informa. A continuación se presentan en forma de tabla los principales componentes.

Cognitivismo	Conductismo
Lo real es lo que está presente a la conciencia	Sólo es real lo observable y medible con procedimientos replicables
El aprendizaje se inicia en el sujeto	El aprendizaje se inicia en el ambiente
El aprendizaje es controlado por el sujeto; control endógeno	El aprendizaje es controlado ambientalmente; control exógeno
Aprender implica modificación de los esquemas cognitivos del sujeto	Aprender implica la modificación de patrones de respuesta en el sujeto
Enseñar implica poner condiciones que estimulen y faciliten la actividad del sujeto	Enseñar implica disponer el medio en término de los estímulos y recompensas a que será expuesto el sujeto
La relación de colaboración facilita e incrementa la calidad del aprendizaje	La relación de competencia fomenta el aprendizaje de patrones exitosos

2.4. Una visión relacionada.

Sin intentar establecer una relación de continuidad entre los modelos teoréticos filosóficos y psicológicos, podemos pensar en los énfasis relativos de cada posición, asociando el modelo del criticismo kantiano con los planteamientos cognitivos, y el modelo positivista con el conductista (Ver gráfica).

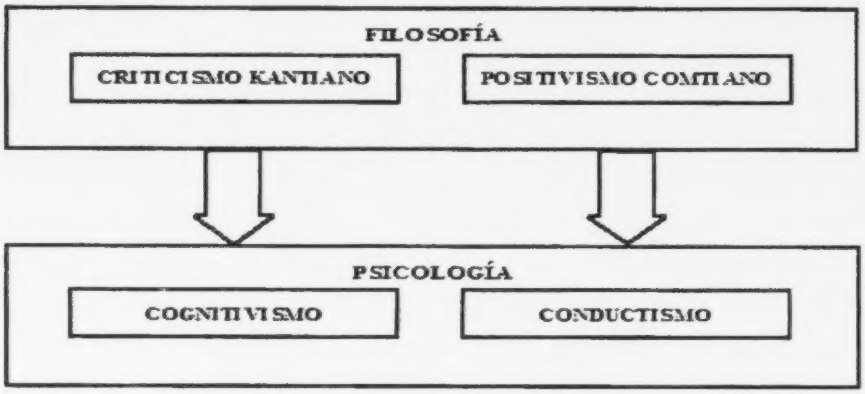


Ilustración 1. Referentes Filosófico y Psicológico

Se aprecia cómo los planteamientos del criticismo kantiano enfatizando la actividad de la razón pura se relacionan con la propuesta del cognitivismo, que también enfatiza los procesos internos de la conciencia. Por su parte, los planteamientos del positivismo comptiano se relacionan de buena forma con la propuesta del conductismo.

Lo anterior es comprensible pues ambas posiciones filosóficas corresponden no sólo a antropologías sino que a epistemologías, a visiones acerca del origen, naturaleza, dinámica e impacto del conocimiento científico, tanto como a ontologías, es decir, especiales concepciones acerca de la realidad.

3. Aprendizaje.

Frente al tema del "aprendizaje" hay una variedad de opiniones en los sujetos entrevistados, las cuales pueden asociarse a alguno de los modelos examinados. En los siguientes párrafos se examinan las respuestas a la luz del marco integrador que se muestra en la ilustración siguiente.

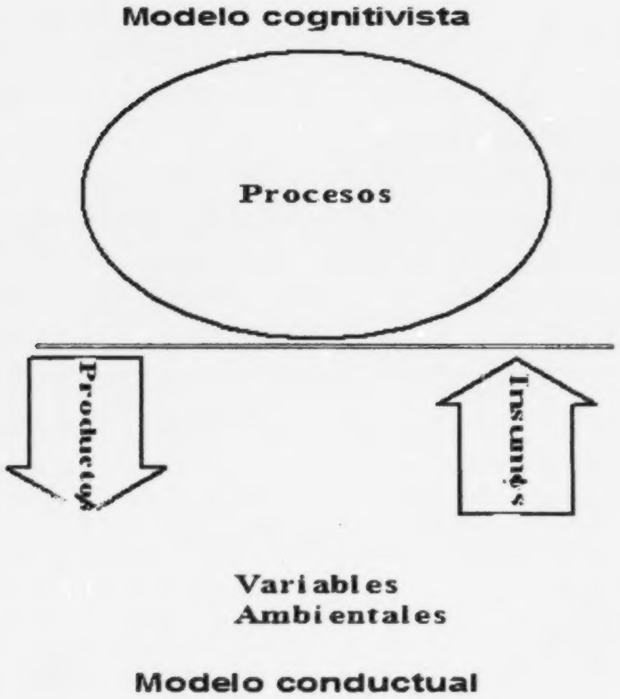


Ilustración 2. Dos modelos de aprendizaje.

3.1. El concepto de aprendizaje en el discurso docente.

Concepto conductual.

Una primera concepción del aprendizaje es la denominada "tradicional", identificada con lo memorístico, mecánico y repetitivo. Lo que es rechazado por los docentes, al menos en el plano del discurso. Como se verá, en ella aparecen algunos rasgos notorios del modelo conductual, donde es importante el concepto de cambio o paso de un estado inicial A (ignorancia) a un estado final B (aprendizaje). La Ilustración 3 sintetiza los aportes en este plano.

El aprendizaje es caracterizado como: "*los estudiantes memorizan cosas, memorizan conceptos, definiciones que no siempre están bien concluidas*" (E01, 30-31); se trata entonces de un aprendizaje que consiste "*en memorizar y transformarse en diccionarios o en un glosario de términos*" (E01, 46-48). Esta manera de concebir el aprendizaje se asocia muy bien con el concepto que Paulo Freire denomina "educación bancaria", en la cual el docente "deposita" una cantidad de información en la mente del estudiante, para luego "retirarla".

Se entiende así el aprendizaje bajo el patrón de la transmisión de contenidos desde alguien que sabe a alguien que no sabe: el aprendizaje se produce cuando alguien "*transmite a alguien que desconoce sobre el tema que se está tratando*" (E02, 20-21), de manera que "*el alumno es el que en el fondo va a poder asimilar este conocimiento de forma tal de que Él lo pueda usar hacerlo propio*" (E02, 35-38). En términos más sutiles, también se encuentra la expresión (que) "*tenga a su disposición o pueda él manejar, conocer, con cierta facilidad una serie de disciplinas eh... que lo habilitan para posteriormente ejercer una profesión, eso sería aprender universitariamente una carrera una profesión y dentro, ahora ya, más específico todavía de mi propio quehacer*" (E04, 34-41). También se lo expresa como "*aprender, aprendizaje es aprender, eso tiene que ver con adquirir conocimiento*". (E05, 18-20).

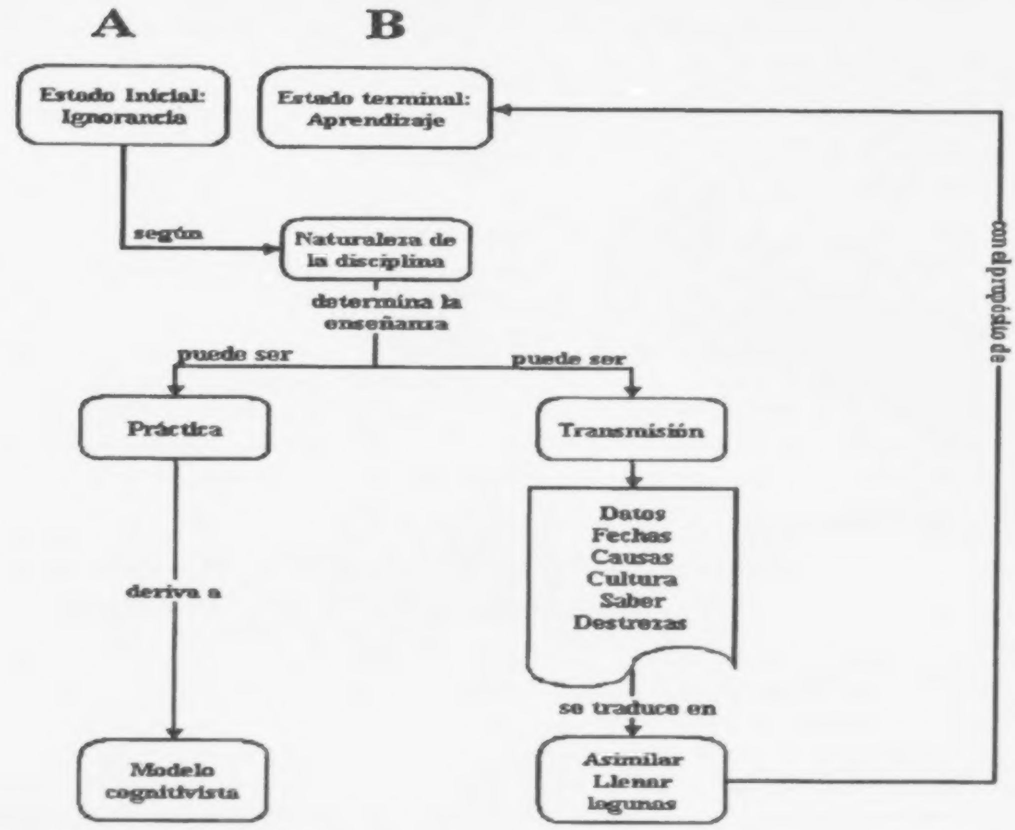


Ilustración 3. Modelo conductual del aprendizaje

Esta concepción se asocia con aprendizaje memorístico o "tradicional"; sin embargo, no es una asociación gratuita, sino que se la concibe relacionada íntimamente con la naturaleza de la disciplina que se enseña. De esta forma, un docente indica que "*estoy en estos momentos dictando química orgánica, primer semestre, porque es más memorización que trabajo matemático*" (E02, 436-438). Aquí se advierte con toda claridad que la química orgánica se asocia a un trabajo de memorización, mientras que la matemática aparece como "trabajo", es decir como una actividad de parte del estudiante, superior a la mera repetición. Concebir el aprendizaje de esta manera, como un intercambio de información (generalmente en forma de pequeñas unidades de datos o algoritmos), tiene claras implicancias sobre la enseñanza.

Desde la óptica propiamente docente, se trata entonces de disponer de conocimientos "*que los alumnos los asimilen, puedan trabajar con ellos y no queden con lagunas insalvables que después ellos en su vida profesional les haga pasar una gran vergüenza*" (E08, 483-486).

Si bien el concepto de contenido suele referir al currículo, es necesario precisar que el aprendizaje no se realiza en el vacío sino que sobre cuerpos de conocimiento bien delimitados, que son conocidos como curriculum o programas de estudios. Los contenidos asumen diversas expresiones en la terminología utilizada por los docentes entrevistados: "*aprender no sólo los conceptos elementales, básicos, ... sino también... las destrezas*" (E08, 25-28) de manera que los alumnos los "*asimilen, puedan trabajar con ellos y no queden lagunas insalvables*" (E08, 483-484) en cuanto a "*materia, contenido o saber*" (E09, 22-23), así como "*ciencias básicas* [y]... *prácticas tecnológicas*" (E10, 26-27); en general, se entiende que se trata de "*información sobre diversos temas, tópicos, disciplinas*" (E11, 22-23).

El resultado de este enfoque es que "*podemos tratar de transformar entre comillas a estudiantes eruditos pero no estudiantes que sepan*" (E01, 61-63). Se hace una distinción entre "saber" y "erudición", según se explora en párrafos a continuación. Esta forma de concebir el aprendizaje y la enseñanza no es algo nuevo ni propio del sistema universitario sino que la "*educación predominante en Latinoamérica y en Chile hasta hace poco, ha sido la de una docencia de muy embotellar el conocimiento en la cabeza de los alumnos*" (E15, 249-252).

En síntesis, encontramos que entre los docentes una concepción de aprendizaje se asocia al modelo conductual en los siguientes términos:

Elementos del modelo	**Hallazgos**
Sólo es real lo observable y medible con procedimientos replicables	El aprendizaje es real cuando el sujeto reproduce lo aprendido

El aprendizaje se inicia en el ambiente	Los contenidos son entregados desde fuera al sujeto que aprende
El aprendizaje es controlado ambientalmente; control exógeno	El control radica en el docente, no en el alumno
Aprender implica la modificación de patrones de respuesta en el sujeto	Antes no sabía, ahora sabe.
La relación de competencia fomenta el aprendizaje de patrones exitosos	No hay referencia alguna a la colaboración ni cooperación

Concepción cognitiva.

Tradición piagetana.

Las expresiones que asociamos a la concepción cognitiva del aprendizaje son aquellas que hacen referencia a lo que sucede en la "mente" del aprendiz más que los productos que éste pueda emitir hacia un observador externo. Esta concepción está diagramada en Ilustración 4.

Encontramos una metáfora que compara el aprendizaje con el almacenamiento y procesamiento de datos. De hecho, se reconoce que se trata de un proceso mental, en el cual suceden intercambios y utilizaciones. Se indica que el aprendizaje "*es un proceso mental en el cual se cambia un stock de conocimientos que tenemos o la forma como procesamos ese conocimiento, eso que uno tiene en su cabeza uno tiene una serie de datos guardados y esos datos uno los utiliza, los procesa en cierta forma para obtener resultados, luego cuando uno quiere tomar decisiones, entonces el aprendizaje puede cambiar los datos de base de este proceso de toma de decisiones o cambia los procesos por los cuales tomamos decisiones*". (E15, 19-31). Lo notable de esta expresión, más allá de la metáfora del procesador de datos, radica en que hace referencia a la necesidad de que el sujeto reelabore, reestructure sus contenidos de conciencia, con el propósito de tomar decisiones.

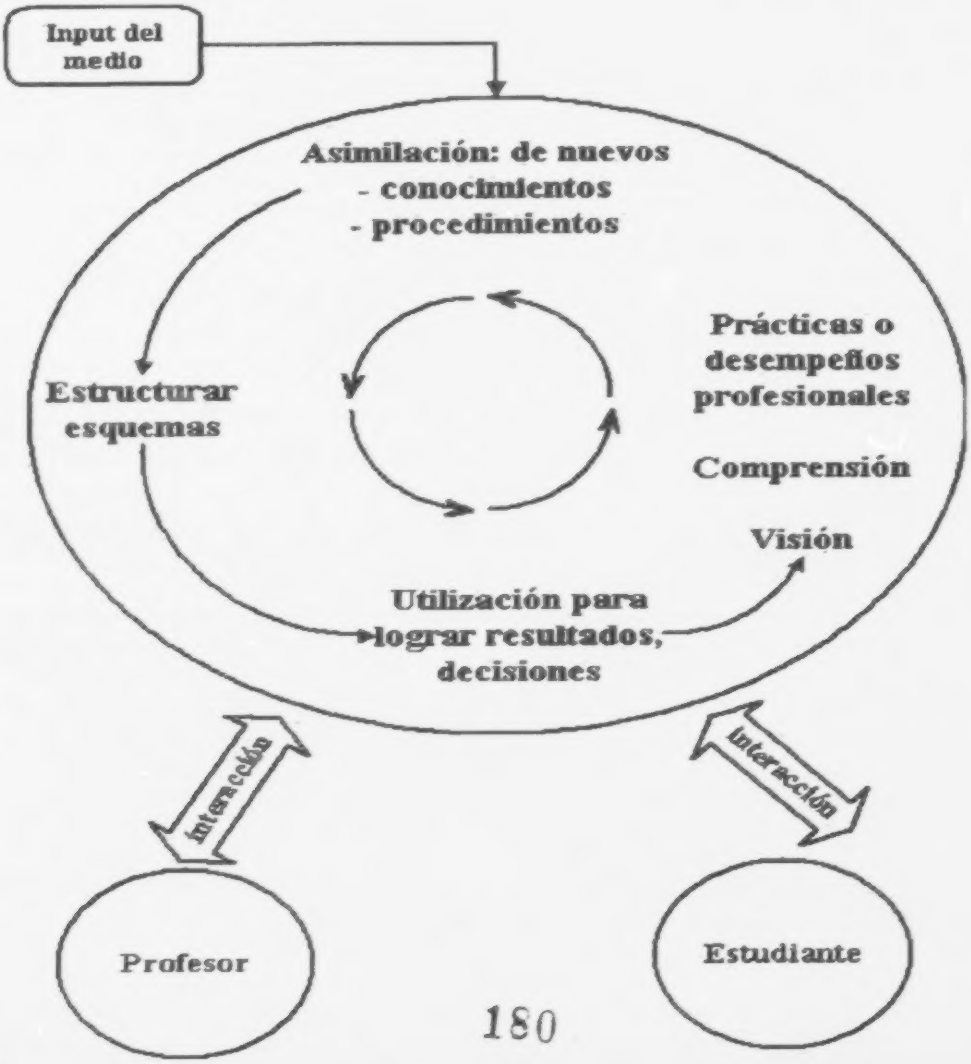

Input del medio
Asimilación: de nuevos
- conocimientos
- procedimientos
Estructurar esquemas
Prácticas o desempeños profesionales
Comprensión
Visión
Utilización para lograr resultados, decisiones
Interacción
Interacción
Profesor
Estudiante

Ilustración 4. Modelo cognitivista

La imagen del procesador de datos es expuesta por un docente que indica que "*mí el aprendizaje es un proceso en el que un grupo de personas, en este caso los estudiantes, reciben procesa e integra información sobre diferentes temas, tópicos, disciplinas pero las tres cosas: recibe, procesa e integra, eso yo diría que son las partes básicas del proceso de aprendizaje*" (E11, 19-26). Los componentes del aprendizaje asimilación y acomodación aparecen claramente en los tres verbos: "recibe, procesa e integra" utilizados por el entrevistado. Ciertamente la actividad intelectual del sujeto constituye el eje de esta concepción del aprender.

Otro docente lo expresa en términos más técnicos señalando respecto del aprendizaje que "*Yo la entiendo como una reestructuración de los esquemas mentales conceptuales que tiene el sujeto, eso para mi es el aprendizaje*". (E14, 17-20).

Variante sociocognitiva.

Frente a la noción "tradicional" que visualiza al estudiante como receptor y repetidor de unidades de información o aplicador de algoritmos, o bien como un procesador individual de información, se plantea un nuevo concepto, en que la interacción del estudiante parece ser un eje fundamental. Por ello se dice que "*lo que más importa es que el estudiante logre discutir cosas, discutir conceptos, discutir estrategias, discutir formas de que esos conceptos se transforman en realidad más que en memorizar*" (E01, 42-46). El estudiante es concebido entonces como un sujeto que no solamente está activo al nivel "mental", en la interioridad de su conciencia, sino que también es activo en la interacción social, particularmente en la relación profesor-alumno.

Esta manera de concebir el aprender impacta en la estructura de relaciones, con lo cual al "*alumno a mi juicio habría que darle el máximo de confianza para que venga al profesor para que él interrogue al profesor, le pida lo que quiere, le pida que lo ayude, pero no tampoco que el profesor sea un esclavo del estudiante: 'mire no haga nada sin que yo esté presente, no haga nada sin que yo lo conduzca, sin que yo lo dirija'*..." (E01, 100-108). Aún cuando insiste en la actividad del alumno, la expresión mantiene la asimetría de la relación docente-discente, en que el aprendiz aparece siempre como dependiente del docente.

Aprendizaje como significación.

En la tradición cognitivista el aprendizaje se produce cuando el aprendiz es capaz de integrar significativamente nuevas unidades de información (en el sentido más lato del término) en sus estructuras o esquemas cognitivos. El punto crucial radica en la significación; de estar ausente ésta, no habrá aprendizaje auténtico ni duradero. Por ello es crucial dotar de significado el aprendizaje: éste es aprendizaje valioso cuando es apropiado para dar sentido y significación a la propia existencia. Un párrafo de un docente lo expresa diciendo que "*nosotros entendemos por aprender ...eh... la constitución de una consciencia en esos estudiantes, sobre el mundo en el que están viviendo, sobre ciertos conceptos, ciertos valores o sobre el mundo que están viviendo, que no se traduce necesariamente en un, en una técnica que vayan a aplicar ellos, sino que se traduce en una actitud frente a la vida y junto con esa consciencia, que es una palabra que me gusta mucho, tomar consciencia, tener consciencia, tener consciencia histórica, tener consciencia estética, que son una de las cosas que a mi me parecen interesantes. Junto con eso yo diría ...eh... saber leer, en el sentido general, textos y la realidad, es decir entender los signos de la realidad, no ir por la vida torpemente sino entendiendo cuáles son los caminos que se abren a mí en la vida, en la propia realidad del mundo, o sea, nosotros por aprender, o sea le pedimos al final de un semestre sería eso, si efectivamente un alumno puede, puede tener para sí esos niveles*" (E04, 65-90).

Así lo importante es que el estudiante universitario no sólo aprende una profesión sino que "*aprende una manera particular de entender el mundo*" (E07, 50-51). Tener conciencia de la historia tener conciencia del mundo, tener conciencia estética, son todas expresiones que se refieren al estar presentes en la historia, el mundo, el arte, pero también a que la historia, el mundo, el arte, estén presentes en la mente y conciencia del sujeto. De esta forma, podrá "*leer ... textos y la realidad*"; más sintéticamente, sólo leer textualizando lo que se ofrece a su conciencia. Todo, incluyendo al propio sujeto, se presenta como un pre-texto sobre el cual el mismo sujeto podrá textualizar y re-textualizar, en un proceso de significación y re-significación permanente de su propia experiencia. Eso es aprender.

Desde otra perspectiva, también intelectual y compleja, el aprendizaje es visto como establecer distinciones no sólo entre los objetos sino distinciones acerca de las distinciones. Alcanzar este carácter doblemente reflexivo de la actividad intelectual representaría un aprendizaje importante. Por ello, un docente pide que "*vayan desarrollando la habilidad de distinguir lo que es importante de lo que no es importante, lo cual implica una serie de cosas, lo que implica que*" ... sepan "*no sólo cómo funciona, quiere decir no sólo cómo funciona una regla y cómo uno puede distinguir lo que es importante de lo que no es importante. Básicamente un estudiante que sabe distinguir lo que es importante de lo que no es importante todo lo demás viene por añadidura y es capaz de aprender él todos los contenidos que sean estudiados, pero eso es, cómo distinguir lo que es importante de lo que no es importante y como aplicar esa distinción a la distinción misma, o sea, como saber si distinguir entre importante o no importante es en sí mismo importante o no importante. Esta segunda definición es muy importante, es crucial*" (E07, 20-42). Nuevamente encontramos la fuerte connotación al dotar de sentido, en este caso expresado como distinguir lo importante de lo no importante. Un sujeto que solamente aprende en el marco del insumir y repetir no será capaz de establecer distinciones y, menos aún, de otorgar sentido a las cosas.

Por consiguiente, aprender "*no es solamente adquirir conocimiento en parte, es adquirir destrezas, es adquirir habilidades, es adquirir conductas, ..eh... Finalmente es casi como... a ver, digamos... es como adoptar la cultura, en el cual tiene muy variados aspectos dentro de eso, pero que está más o menos englobado en lo que le dije en un comienzo, es un cambio de conducta*". (E06, 18-26). En este texto aparece el término "conducta" que no necesariamente puede tomarse en su plena significación técnica. Entonces en el aprender "*se trata no sólo que la persona sea capaz de repetir, o sea, de reproducir las categorías de importancia o no importancia que recibió*" (E07, 43-46). El punto es que la persona sea capaz de elaborar construcciones acerca de la porción de la realidad sobre la cual está dirigiendo su propia formación, con la ayuda de los profesores y de todo el aparato institucional que supuestamente ha sido diseñado y es operado para facilitarle esta tarea.

Elementos del Modelo	**Hallazgos**
Lo real es lo que está presente a la conciencia	Los contenidos de conciencia (aprendizaje) son reales
El aprendizaje se inicia en el sujeto	El sujeto es quien aprende, no se obligar el aprendizaje
El aprendizaje es controlado por el sujeto; control endógeno	El estudiante es quien aprende
Aprender implica modificación de los esquemas cognitivos del sujeto	El estudiante aprende cuando incorpora y modifica sus esquemas
Enseñar implica poner condiciones que estimulen y faciliten la actividad del sujeto	La labor del docente es "entregar herramientas" para que el alumno aprenda
La relación de colaboración facilita e incrementa la calidad del aprendizaje	Se afirma la importancia de la relación pero sólo una relación de dependencia profesor-alumno

3.2. La visión del alumno como sujeto de aprendizaje.

Se examina la visión del estudiante en cuanto es considerado como un sujeto de cambio. Al decir "sujeto" se le entiende en tres sentidos diferentes: sujeto de cambio, como "paciente" del mismo; sujeto de cambios, como lugar donde se producen; y sujeto del cambio, cuando es el mismo estudiante quien inicia, monitorea, ejecuta y evalúa sus procesos de aprendizaje.

Las expresiones de los docentes focalizan en grados diversos el concepto de "sujeto", manteniendo una tensión entre la deseable autonomía del estudiante y una dependencia del profesor. La Ilustración 5 intenta plasmar gráficamente los conceptos.

El estudiante como un sujeto de cambio.

El aprendizaje implica un cambio, una "*transformación que logra el estudiante en algunas áreas,...*" (E01, 53-54). En palabras de otro docente, "*aprender es, sería, pasar de un estado en que yo no sé hacer algo o no conozco algo, a un estado en el que yo puedo hacer ese algo o ya lo conozco, lo domino*" (E04, 23-27). También se entiende el cambio como un pasar de lo que no había a lo que hay, de lo desconocido a lo conocido, de la ignorancia al saber. "*Es la capacidad de hacer propio el conocimiento, es la capacidad de reconocer aquello que era desconocido y hacerlo conocido, es pasar de un estado de desconocimiento de una materia, contenido o saber, a un estado de conocimiento de esa materia, contenido o saber*". (E09, 18-25).

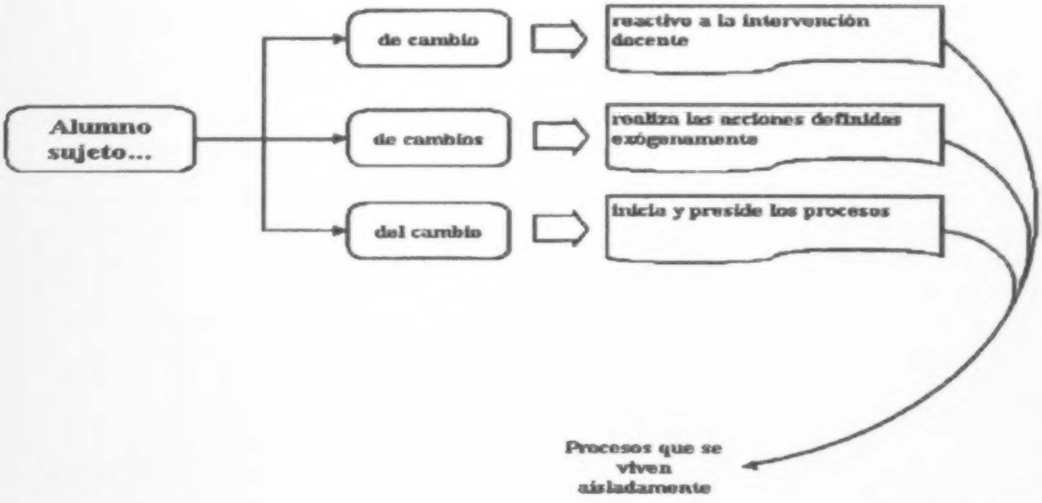

Alumno sujeto...
de cambio
de cambios
del cambio
reactivo a la intervención docente
realiza las acciones definidas exógenamente
inicia y preside los procesos
Procesos que se viven aisladamente

Ilustración 5. Tres visiones del estudiante como sujeto de aprendizaje

El cambio entonces puede ser concebido como un pasaje o variación en el estado de un sujeto. De una situación S se pasa ahora a una situación S+x, donde "x" representa el incremental atribuible a los procesos de aprendizaje. La lectura de la expresión S+x debe precaviéndose de no implicar que x es un monto determinable (tentación asaz presente en los procesos evaluativos o de examen), sino que suele ser un poco más difuso e inasible. Podríamos ocupar aquí la analogía del "efecto de halo", interpretando entonces que el contenido de "x" es menos preciso, comprendiendo también otras dimensiones del sujeto como, por ejemplo, la afectividad, las creencias y adhesiones (éticas, religiosas, políticas), su autoestima, entre otras.

Al ser visto el estudiante como un sujeto activo e involucrado en el aprender, puede pedírsele que actúe y realice algún conjunto de operaciones como "*que él interrogue al profesor, le pida lo que quiere, le pida que lo ayude*" (E01, 102-104), con las prevenciones anotadas más arriba.

En otra expresión, un docente señala que "*yo lo defino como un proceso mediante el cual un grupo de alumnos, con la guía de un profesor, aprenden una materia, lo que no significa que aprendan solamente las clases del profesor sino con la guía del profesor aprenden, con las clases y los trabajos que los mismos realizan, por los laboratorios y por las lecturas. Yo no creo que aprendizaje sea asistir a una clase, aprendizaje es mucho más que asistir a una clase*" (E03, 22-33). SE ve una indicación sobre la importancia que el alumno sea activo en el proceso de aprender. También es destacable que se refiera a un grupo de alumnos. Si bien no alude a trabajo cooperativo ni colaborativo, sin embargo, la aparición del término "grupo" puede conducir a una conciencia de la heterogeneidad de los estudiantes. No obstante, en esta expresión no se resuelve cómo se concibe el "*aprenden una materia*": esta puede ser leída tanto en clave conductista como cognitivista.

De todas las expresiones anteriores puede inferirse entonces que el estudiante es considerado como un sujeto, pero no proactivo en el cambio, sino que más bien reactivo a los componentes ambientales (docencia) a los que se encuentra expuesto sistemáticamente.

El estudiante como un sujeto de cambios.

La expectativa sobre el aprendizaje es otra. La docencia, por su parte, aporta al estudiante para que éste no sólo pueda aprender conceptos sino también destrezas, incluso la posibilidad de que ellos mismos proyecten sus propios aprendizajes. "*...un proceso a través de los cuales existe una interrelación entre un alumno y el profesor, el profesor qué es lo que pretende a través de este proceso entregando las herramientas necesarias para que los alumnos aprendan no sólo los conceptos elementales, básicos que le sirvan para desenvolverse, sino también adquieran las destrezas para que por sí mismos sean capaces de buscar los objetivos de la disciplina, los objetivos específicos de la disciplina que uno está enseñando*" (E08, 19-32).

Por cierto, no queda claro de la expresión anterior qué se entiende por "objetivo". Es posible que exista una connotación diferente a los tradicionales "objetivos educacionales" o "instruccionales"; se puede percibir una sugerencia relativa a cuáles son los propósitos cognoscitivos de una disciplina determinada. En otras palabras, en la expresión señalada el concepto "objetivo" sería más bien de orden epistémico antes que pedagógico o didáctico. A pesar de todo, se reconoce que el aprendizaje implica una acción determinada, en y a través de la cual se produce el fenómeno de aprender. La tarea del docente aparece vinculada a un proceso de ayuda, de "*entregar las herramientas necesarias*" pero sin convertirse en el actor principal, rol que está reservado al alumno.

Esto es especialmente destacado cuando se trata de la formación tecnológica o profesional. En el caso de Agronomía, se indica que se trata de aprender "*prácticas ya tecnológicas, por lo cual no involucra sólo obviamente esas transferencias sino que el mismo estudiante pueda practicarlas antes que, practicarlas para que pueda ver el concepto al menos así lo vemos en agronomía*" (E10, 27-33).

Similarmente, para la formación en la carrera de Tecnología Médica se indica que "*entonces, y los tenemos un año en esas condiciones son dos semestres con igual número de horas y el alumno, como le digo, ¿qué es lo que uno pretende?: que me sepa hacer un perfil lipídico. Eso, yo diría, es una enseñanza, eso es el aprendizaje de ese alumno, que puede ser distinto a otro -para el abogado puede ser totalmente diferente, a lo mejor para el abogado es bueno hablar que pueda expresarse en público-, en cambio este alumno tiene que saber hacer sus cosas con las manitos*" (E12, 91-102).

El estudiante como sujeto del cambio.

Finalmente, aparece una expresión que insiste con claridad sobre la necesidad de que el estudiante se involucre activamente como actor central del proceso de aprendizaje, donde él sea el sujeto que produce, no sólo que padece o en el cual se observan los cambios.

No obstante, la expresión aparece como una crítica directa al estilo de enseñanza de la educación media, indicando que "*nunca se les ha hecho discutir por qué, por ejemplo, una situación es mejor o peor, cuándo es más o menos beneficio, cuando o por qué pasó o por qué pudo haber pasado tal cosa, que habría sucedido si hubiera acontecido esto otro*" (E01, 149-155).

3.3. El contenido del aprendizaje.

Se examinan las expresoines de los docentes que se relacionan con los logros o productos del aprendizaje. Estos son calificados de diversas maneras: como contenidos, competencias, dominio de algo, logro de objetivos. La Ilustración 6 presenta esquemáticamente las acepciones.

(a) Aprendizaje como contenidos.
La concepción conductual se asocia con el aprendizaje memorístico pero no es gratuita, sino relacionada íntimamente con la naturaleza de la disciplina que se enseña. De esta forma, un docente indica que "*estoy en estos mnomentos dictando química orgánica, primer semestre, porque es más memorización que trabajo matemático*" (E02, 436-438). Se advierte que la materia

se asocia a un trabajo de memorización: lo que tiene que hacer el estudiante es "saber" la materia, es ser capaz de repetirla. En cambio, la matemática aparece como "trabajo", es una actividad del estudiante superior a la mera repetición.

(b) Aprendizaje como logro de competencias .
Las competencias son concebidas como capacidades para operar, ya sea en el plano intelectual como en el manual. Se señala que lo importante es que "*lo haga, yo creo que adquirir esa capacidad de análisis es lo que más le*" (E01, 164-165) ayudará al estudiante a lograr el éxito, "*más que memorizar cosas, conceptos, aprenda a aplicarlos*" (E01, 280-281). Esta competencia intelectual tiene que ver no sólo con que el estudiante sea "*capaz de aplicar esas categorías a las categorías mismas, o sea, no es sólo que el estudiante de derecho aprenda sólo derecho aprende una manera particular de entender el mundo, por eso digo que no sólo distinguir lo importante de lo no importante, sino ser capaz de distinguir que las categorías que usa también están sujetas a la definición*", (E07, 47-55), lo cual ciertamente hace una demanda sobre la capacidad de reflexión o el logro de aprendizajes reflexivos por parte de los estudiantes.

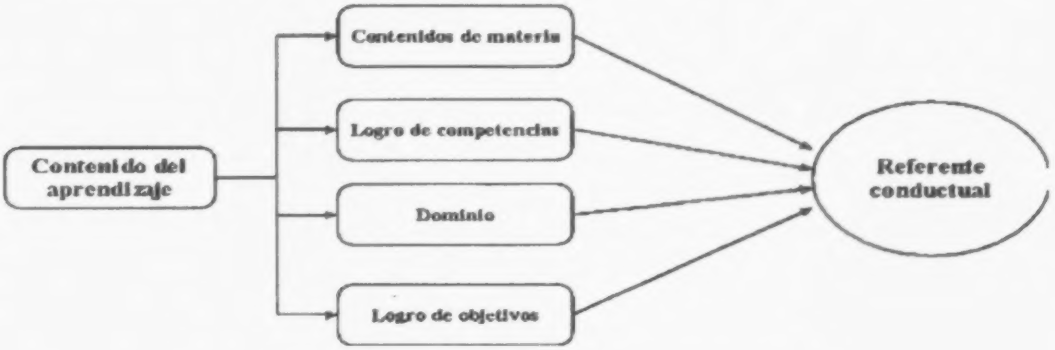


Ilustración 6. Acepciones del contenido del aprendizaje

También las competencias se conciben en referencia a la profesión, entendiendo que lo importante es que los estudiantes "*adquieran tales o cuales características profesionales*" (E01, 240-241); específicamente, "*la transferencia de conocimiento y en el área de la agronomia es muy aplicado a las prácticas de técnicas de cultivo o traspasar la ciencia básica a las practicas ya tecnológicas*" (E10, 23-27)

(c) Aprendizaje como dominio.
El dominio es concebido latamente como una capacidad de operar sobre un conjunto dado, que puede estar compuesto por entidades cognoscitivas o bien materiales. Se dice que "*para mí el alumno ha aprendido ha aprendido cuando, en términos vulgares, uno lo da vueltas hacia el revés o hacia el derecho y le dice aplique aquí, critique aquí, analice aquí y él es capaz de hacer eso*". (E01, 55-59). También se expresa como "*aprender es, seria, pasar de un estado en que yo no se hacer algo o no conozco algo a un estado en el que yo puedo hacer ese algo o ya lo conozco, lo domino*" (E04, 23-27).

La noción de dominio se diferencia del clásico concepto de "mastery" desarrollado por Carroll, Bloom y otros de la Escuela de Educación de Chicago, a fines de los años 60. Ese concepto estaba concebido más en términos de la adquisición de un determinado contenido o destreza, dependiente del esfuerzo o tiempo activo en la tarea de aprender. Sin embargo, la noción de "mastery" no escapa al paradigma conductista. En cambio, en el caso de los docentes universitarios el concepto de dominio se acerca más al modelo cognitivo, teniendo que ver con las habilidades intelectuales y procedimentales que el sujeto logra ejecutar sobre un conjunto delimitado del saber.

(d) Aprendizaje como logro de objetivos.
El término "objetivos" es utilizado directamente por los docentes, sin mayor especificación. De acuerdo a las características de la cultura local, se los entiende como propósitos que se plantean en las diferentes asignaturas. De esta forma se asume que un estudiante aprende cuando hay "*elementos que lo conducen a creer que ha logrado objetivos*" (E01, 38-39), "*a esos objetivos vamos viendo si el alumno los está consiguiendo o no a lo largo de la asignatura*" (E12, 24-26). Los objetivos pueden ser de diferente naturaleza o alcance; entre ellos los de "*carácter profesional*" (E01, 228).

No se encuentran, sin embargo, referencias acerca de la genética de los objetivos, apareciendo como "dados" en una constelación representacional que podemos denominar el currículum.

3.4. Implicancias de un buen aprendizaje.

Lograr aprendizajes de buena calidad no es sólo una cuestión técnica, asociada al ejercicio de la actividad docente y discente, sino que tiene un fuerte componente ético, desde la óptica docente. De esta forma, un académico señala que "*para un dentista debe ser fundamental aprender a realizar bien una operación, porque de otra manera corre el riesgo de infecciones y pone en peligro la vida humana y para un abogado tiene que saber aprender ciertos aspectos del código porque podría condenar, digamos a cadena a un inocente, condenarlo por un delito que no cometió si no hace bien su profesión, si no aprendió en la universidad a defender un inocente pues corre el riesgo de involucrar una vida humana y el caso de que, cuando habia pena de muerte, condenarlo a muerte por ejemplo*" (E04, 47-61).

3.5. Los medios y recursos del aprendizaje.

La tecnología, especialmente la TIC, es escasamente referida por los docentes. De hecho existe sólo una referencia explícita en que se hace un paralelo hablando de "autoaprendizaje" (a pesar de lo redundante que puede ser el concepto, como si fuera posible un hetero-aprendizaje), expresando un grado de desconfianza hacia la tecnología. Se ha "*escuchado hablar del auto aprendizaje, el auto aprendizaje se asimila a unos computadores y cursos en la red, entonces el estudiante va y aprende y eso no es aprendizaje un estudiante se puede meter en la red y ver que bonitos son los monitos que colocaron, pero para que haya verdadero autoaprendizaje tiene que haber una participación del estudiante por meterse por eso que está ahí*" (E11, 441-450). La visión que tiene este docente del potencial pedagógico de las TIC, es rudimentaria, asimilando el uso de la tecnología para la instrucción a la sobresimplificación de los conceptos y al peligro de perderse en los abalorios, en los adornos de los diseños que suelen encontrarse en las páginas web.

Otra enfoque de la tecnología es mediante su incorporación en los patrones de aprendizaje de los estudiantes, entendiendo que aprender la tecnología es un contenido fundamental para los tiempos que corren y los futuros. De hecho, un docente expresa que se requiere preparar profesionales para los nuevos tiempos, en que las operaciones de rutina estarán por lo general automatizadas, debiendo en cambio enfrentar problemas emergentes, situaciones nuevas, de excepción, frente a las cuales no valdrán recetas sino que lo más importante será la creatividad y la capacidad crítica y reflexiva para enfrentarlas. De esta forma, se plantea que "*generar aprendizaje, en la nueva sociedad, digamos, que se está formando, cada vez más nos veremos enfrentados a menos situaciones rutinarias y a más situaciones de excepción, eso es porque bueno la tecnología está automatizando muchos procesos por lo cual todo lo que se pueda hacer rutinario puede ser automatizado y lo que puede ser automatizado lo van a hacer algún tipo de máquina, lo cual la gente cada vez más va a tener que manejar situaciones de excepción; eso implica que no siempre se puede aplicar una receta y entonces hay que conocer diferentes alternativas de ver un problema y utilizar el criterio para una alternativa en una situación en particular*" (E15, 256-273). Como se advierte de la expresión, la tecnología es concebida no como un auxilio para aprender, sino como una herramienta que el futuro profesional debe aprender a utilizar de tal manera que pueda dedicar su atención y creatividad a las situaciones de excepción, a aquellas que desafían las rutinas algoritmizadas en los dispositivos de control digital.

De esta forma, podemos decir que en la opinión de los entrevistados el rol de la tecnología y específicamente de las TIC es prácticamente nulo como dispositivo que colabore al aprendizaje de los estudiantes.

3.6. Síntesis.

Los principales hitos que han podido determinarse en el análisis de la concepción de los docentes respecto del aprendizaje son los siguientes:

- El aprendizaje es entendido alternativamente bajo el paradigma conductual y bajo el cognitivista;
- El tipo de aprendizaje esperado tiene que ver con la visión epistémica que tienen los docentes de su propia disciplina;
- La dimensión relacional o interpersonal del aprendizaje se plantea únicamente entre profesor y alumno; no se considera el trabajo colaborativo entre estudiantes;
- La labor del docente se expresa más como "entregar" (contenidos o herramientas) que como promover la autonomía intelectual del estudiantes;
- Hay un énfasis en funciones de comprensión, especialmente respecto de un concepto amplio de cultura o mundo;
- El alumno es concebido como sujeto al que le acontecen los cambios más que como uno que produce cambios;
- El producto del aprendizaje es concebido de diversas formas: como contenidos, como logro de competencias, como dominio, como logro de objetivos;
- Hay una visión descendida de la tecnología, especialmente de las TIC;
- Se expresa un nivel de conciencia respecto de la dimensión ética de la función de enseñar.

4. Enseñanza

Es natural que asociado al aprendizaje, se encuentre la pregunta por la enseñanza. Máxime si la fuente de información son docentes universitarios. Las preguntas que orientan el análisis son ¿qué es la enseñanza? ¿qué sentido le otorgan los docentes? ¿cómo se produce, es decir, cuál es su dinámica interna? ¿cuáles son las estrategias preferidas por los académicos?

Ahora bien, a diferencia del caso del aprendizaje, no se encuentran una o dos teorías de la enseñanza propiamente tales. Lo que se tiene son algunas generalizaciones empíricas y muchos escritos de orden aplicado: cómo enseñar tal cosa o cómo enseñar aquella otra. (Nota 2) Es posible que ello se deba al retraso que tiene la pedagogía como disciplina independiente, consecuencia de haber sido colonizada tantas veces por la psicología, la sociología, la tecnología, y haber quedado vaciada de un discurso propio (Bernstein, 2000; Díaz Vila, 2001).

4.1. Concepto genérico en torno a la enseñanza.

La visión sobre la enseñanza, al parecer, debería ser más clara que la del aprendizaje, pues es más cercana a los actores. No obstante, la primera nota de advertencia la pone un docente que plantea "*Yo creo que en eso estamos bastante rudimentarios*" (E09, 246-247). Esta "rudimentariedad" se revela fundamentalmente en carácter más que nada empírico que reside en el origen de las concepciones acerca de la enseñanza.

A continuación revisamos los principales tópicos, los que se encuentran graficados en la Ilustración 7 de la página siguiente.

(a) Enseñanza como transmisión.

Una de las opiniones más comunes que aparece es la señalar que la enseñanza, en cuanto "*nuestra razón de trabajo es entregar conocimientos básicos, entregar la cultura básica*" (E02, 284- 286). O, en palabras de otro docentes, "*Enseñanza... para mi la enseñanza es la transmisión de lo que ha sido de lo que yo he aprendido en la vida y mi experiencia*". (E03, 37-40). En ambas expresiones el elemento común es "entregar" o "transmitir": realizar un paso desde el profesor al estudiante (E08, 36-37; E08, 193-184; E08, 317; E08, 482; E09, 58). El profesor se mantiene como el actor central del proceso; el alumno es invisible.

Hay quienes se oponen a esta concepción, indicando que si bien el propósito del enseñar es que se produzca el aprendizaje, la relación no es mecánica: "*la docencia tiene el desafío de producir el conocimiento; no lo produce de manera automática, no es un traspaso así como así*" (E05, 50-52).

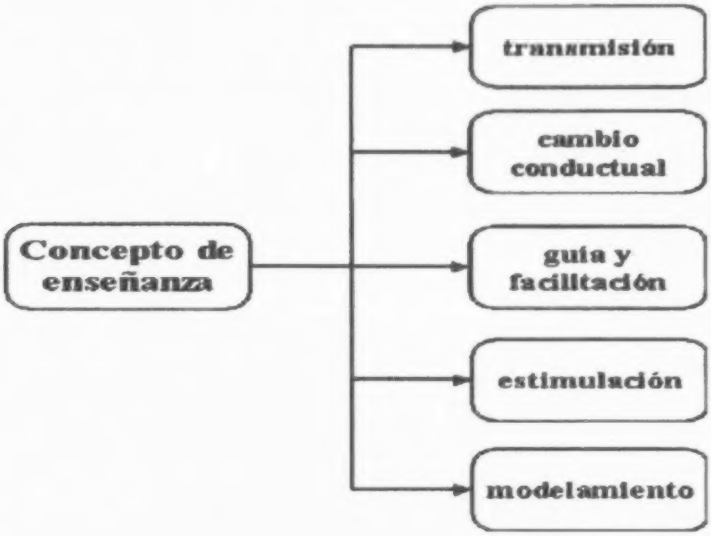


Ilustración 7. Concepciones de la enseñanza

(b) Enseñanza como provocación de cambio conductual.

En otra perspectiva, a la vez que la enseñanza es vista como un deber propio de la tarea académica, el contenido de éste es orientar y transformar a las personas. "*Esa tarea sí debemos hacerla, y estamos obligados a hacerla en la universidad. Es decir, tenemos que ir poco a poco cambiando esas conductas del colegio*" (E09, 164-168) para aproximarse a un modo universitario de aprender.

Aparece en la expresión anterior que la tarea de enseñar se hace equivalente a ir "*cambiando esas conductas del colegio*". Esto tiene dos connotaciones que vale la pena destacar: una, la idea de cambiar conductas, expresión que se asocia directamente con el modelo conductista del aprendizaje y de la enseñanza, según el cual el propósito de toda enseñanza es provocar cambios en las conductas de los estudiantes. La segunda connotación es la diferencia que pretende hacer entre el colegio (la enseñanza media) y la universidad.

(c) Enseñanza como guía y facilitación.
También es posible encontrar visiones alternativas de la enseñanza. Una de ellas, destacada también en el conjunto de las respuestas, dice relación con la concepción de enseñanza como un proceso de guía, en que ésta es realizada por alguien que ya ha hecho los caminos, que ha pasado los procesos y, por consiguiente, puede actuar para reducir los tiempo de búsqueda y de ensayo por parte de los alumnos.

Señala un académico que "*respecto a los estudiantes, yo puedo ayudarlos más fácilmente a hacer el camino a descubrir a que si yo los pusiera solos en una biblioteca a decir mire usted con un mes y veamos que pasó con estas primeras ideas liberales en América Latina, que seguramente sería un trabajo muy largo. En cambio para mi enseñar sería acortarles ese trecho, indicarles dónde están los textos, hacer la síntesis correspondiente ...*" (E04, 112-122). Aquí hallamos el rol docente de estructurar el medio en orden a que los estudiantes produzcan las conductas apropiadas.

También se expresa que la enseñanza consiste en "*guiar al estudiante en el desarrollo del conocimiento, de eso que era desconocido a una etapa de saber*". (E09, 34-36), pasando de un estado (lo desconocido) a otro (lo conocido). Esta noción de guía o conducción se relaciona muy bien con la imagen del "pedagogo" de la antigua Roma, aquel esclavo que conducía al niño por la ciudad y también por los caminos del saber.

Otro tema es el contenido de la propia enseñanza. En las expresiones anotadas anteriormente hallamos que lo que se transmite o entrega no es lo mismo: para uno son los contenidos básicos de una disciplina; para el otro, la propia experiencia vital. Para otro docente, en cambio, "*la enseñanza, ... la enseñanza es, bueno... la transmisión de todos esos factores, en lo posible, o sea, es un proceso muy dinámico en el cual no solamente se transmite el conocimiento sino también se transmiten valores, conductas, ética, además el aprendizaje es estimular la creatividad, el análisis, la crítica, o sea es un proceso complejo*" (E06, 30-39).

La expresión anterior incorpora una dimensión cognitiva tanto como una dimensión ética, lo que en términos de Bernstein podemos denominar el discurso regulativo de la práctica docente. Sin embargo, agrega el docente que además se trata de estimular la creatividad, el análisis, la crítica. Con gran agudeza logra concluir que se trata de "*un proceso complejo*". El punto más crucial radica en cómo conciliar apropiada y equilibradamente una concepción de la enseñanza que combina un eje conductista (transmisión de contenidos y valores) con otro más adentrado en el modelo socio-cognitivo del constructivismo. Esta es una pregunta abierta.

(d) Enseñanza como estimulación.
Una visión diferente es propuesta por otros docentes, para quienes la enseñanza aparece como un proceso mediante el cual se estimula, se proporcionan oportunidades de aprender. Las referencias obtenidas sugieren una visión de procesamiento de datos muy en la perspectiva de la psicología cognitivista, aunque ninguno de los entrevistados posee formación en el área. Así, dice uno de ellos que "*la enseñanza es el esfuerzo o el intento que hace el docente para producir esos cambios (reestructuración de los esquemas mentales conceptuales que tiene el sujeto) en el estudiante*". (E14, 23-27). A pesar de la fuerte referencia psicológica queda claro que la actividad del estudiante de mantiene en segundo plano, siendo más relevante la del docente que será quien produzca los cambios. Como dice otro entrevistado, "*entonces involucra el desafío de producir el conocimiento*". (E05, 29-30)

Una referencia que trae al primer plano la actividad del estudiante indica que "*la enseñanza actualmente, yo la interpretaría como el proceso que permite que los alumnos aprendan, los alumnos cambien su stock de conocimientos y la forma como procesan estos conocimientos, pero ese es un proceso en todo caso facilitador de los que ocurre en la mente de los estudiantes*" (E15, 35-42). La enseñanza entonces es una acción orientada a "permitir" procesos en los estudiantes, ya sea de acumulación o reemplazo de saberes, como también estrategias de procesamiento. Lo notable de esta referencia es que deja en claro que es algo que ocurre en la mente de los estudiantes. No hay profundización sobre qué ocurre en la mente de los estudiantes la que permanece como una "caja negra".

(e) Enseñanza como modelamiento.
El docente puede convertirse en un modelo para el aprendizaje de los estudiantes en la medida en que modele para ellos comportamientos de aprendizaje efectivos y no meramente repetitivos de textos, fórmulas, expresiones. Señala un profesor al respecto que "*la entrega de un conocimiento no es una cosa dura, no es la fórmula, no es... sino que tiene que ver con la forma de enfrentarse a los fenómenos, a los problemas, y eso yo creo que uno lo plantea de una manera novedosa*" (E05, 192-197).

No deja de llamar la atención la última frase que afirma que la enseñanza la plantea de una "*manera novedosa*". Es posible que el docente entienda que hacer cosas novedosas puede ser innovativa, pero también puede suceder que se entienda que el centro de la cuestión radica en lo "novedoso", lo interesante o atractivo de lo que se hace, más que en su valor pedagógico de innovar las prácticas.

4.2. Estrategias de enseñanza.

Las estrategias de enseñanza dicen relación con el tipo de acciones que emprenden o dicen emprender los docentes y la forma de organizarlas. El tema es relevante en el contexto de la universidad y poco a poco se está convirtiendo en un eje de los procesos de reflexión así como de la manera de concebir y evaluar el rol del académico en la institución universitaria.

"*Sí, yo creo que hay más o menos consenso ya en la sociedad chilena, por lo menos en los estamentos universitarios, de comenzar a poner en debate, en la mesa del debate, el tema de la metodología universitaria. Esta misma Universidad de N.N. está preocupada por hacer mucho, está formando comisiones, está trabajando sobre ese tema, no es fácil tener una respuesta inmediata, uno podría evitar en el fondo el asunto dando meramente palabras es decir queremos una clase más interactiva, queremos una clase más humana, queremos una clase más, entonces al final eso se transforma meramente en palabras y no se traduce en un hecho y es un tema muy complejo. Pero yo creo que efectivamente, digamos, mi esperanza es que al menos esta universidad está... está tomando el tema del aula como una cuestión fundamental .. la Universidad de N.N., y esto no lo digo por orgullo, ha hecho algunas reformas y no ha quedado conforme con esas reformas y está volviendo a hacer nuevas reformas dentro de las reformas para... para que justamente la entrega del conocimiento, de las profesiones (¿tú me preguntas de las profesiones?) tiene que ser, tiene que ser de otra manera. Eso no significa bajar las exigencias ni tomar menos pruebas, ni reducir esto o esto otro: es aproximarse de otra manera al conocimiento y eso es difícil de hacer y es complicado.*" (E04, 325-359).

El claro reconocimiento de la importancia del tema de la enseñanza que queda de manifiesto en el extenso párrafo anterior ahorra comentarios. Se percibe una conciencia de los esfuerzos institucionales por mejorar las condiciones docentes. Ciertamente diversos argumentos y razones sostienen esta preocupación institucional por la enseñanza, y aunque no sea este el espacio explícito para plantearlo, es necesario dejar constancia de los ingentes esfuerzos institucionales por el mejoramiento de la enseñanza y, consiguientemente, del aprendizaje de sus estudiantes. Es lo que con sus palabras expresa un docente cuando dice estar "*tratando de buscar el acomodo para ver la mejor forma de que estos chicos, digamos, aprendan*". (E02, 447-449).

Con todo, no siempre se logran estos elevados propósitos; el peso de una formación intelectualmente autoritaria y basada más en la respuesta que en la pregunta transforma a los estudiantes en seres casi completamente modelados y resistentes a una enseñanza que les requiera otro tipo de actividades, actitudes y competencias. "*Yo me he sentido que me he ido perfeccionando, he tratado de hacer las clases lo más entretenido que puedo y veo, igual, choco con una pared dura*" (E05, 234-237). Esta expresión de frustración del docente es inmensamente importante porque trae a la luz una estructura de comportamiento instalada en los estudiantes, que han sido formado dependientes y no autónomos durante doce años de escolaridad.

La presenta en forma diagramática los principales conceptos estratégicos de enseñanza.

(a) La estrategia de la nivelación.
La nivelación es un concepto bastante utilizado en el lenguaje de los docentes, aunque con diversas connotaciones.

En primer lugar, el tema surge cuando existen grupos o cursos con claros desniveles entre ellos. Esto es particularmente en los cursos de primer año, en que los estudiantes se diferencian notablemente sobre las diferentes asignaturas. Como resultado, algunos profesores intentan enfrentar el grupo completo sobre una noción de base estadística: "*generalmente uno trata de enseñar hacia la... hacia la mediana, digamos... no hacia los extremos, y tenemos nosotros que los alumnos muy buenos a veces se aburren, o pierden el interés por la materia. [...] y con pérdida de los alumnos más malos y con pérdida de tiempo de los alumnos mejores*" (E06, 362-366, 375-377). La enseñanza a partir del valor medio (mediana, media aritmética) implica una visión del alumno promedio, abstracto, homogéneo, desprovisto de toda connotación personal y cultural. Los alumnos se hacen transparentes. Sus conocimientos previos no son considerados.

Otra dimensión se refiere a la preocupación porque los estudiantes dispongan de los conocimientos, competencias y destrezas requeridos para poder iniciar el estudio en un curso determinado. Una de las disciplinas que mayores dificultades presenta es la matemática; como dice un docente, "*de hecho en matemática se hicieron estos cursos de nivelación a mediados de año con esa finalidad de poder rescatar a la gente*" (E02, 334-337).

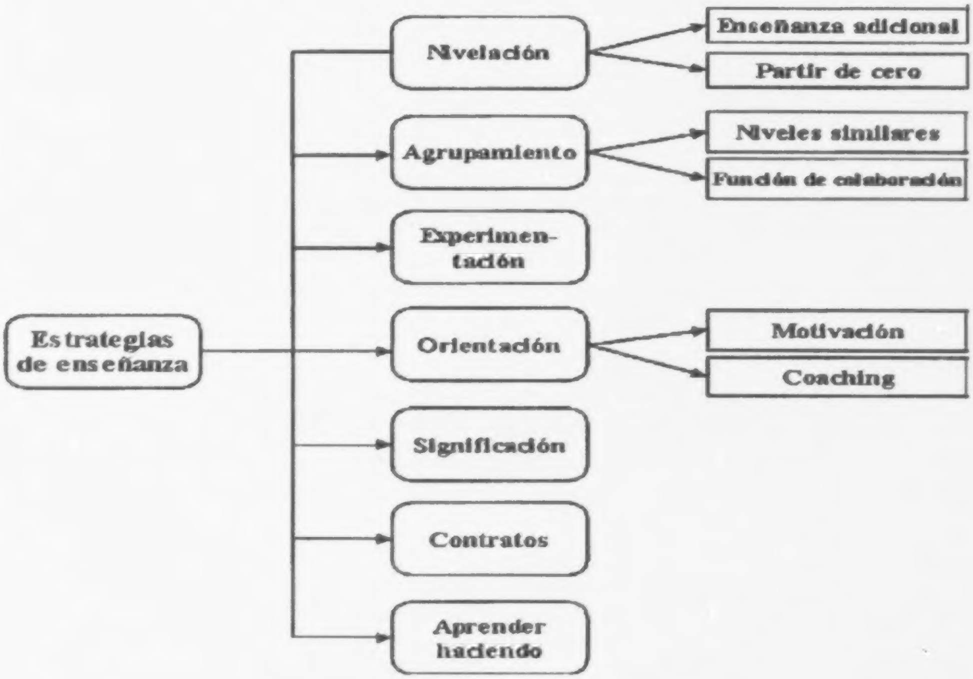


Ilustración 8. Estrategias de enseñanza

El sentido del "rescate" es ilustrativo para representar el espíritu que inspira a estos procesos: la intención de restacar o salvar al que está casi perdido, al estudiante que está a punto de abandonarlo todo, que claramente va al fracaso. En una expresión más extensa, un académico expresa que "*Yo en general no soy partidario de nivelar hacia abajo, soy partidario de apoyarlos, de ayudarlos a ellos; no decir: mire, hay unos tipos que se están quedando, bajemos el nivel de todos. No, yo creo que soy muy partidario de los cursos de nivelación*" (E03, 298-304). No obstante, esta posición no es compartida, existiendo quien asume que la nivelación debe hacerse hacia abajo, hipotetizando la ignorancia como punto de partida: "*me he dado cuenta que.. que hay que partir de cero, no hay que asumir nada*" (E05, 185-187). Esta última posición nuevamente pone el tema de la transparencia de los estudiantes, el desconocimiento de sus conocimientos previos y, consiguientemente, la fijación de un punto "cero" que pasa a ser completamente arbitrario.

Una segunda acepción con la que se habla de nivelación tiene que ver con la expectativa de "*uniformar*" (E02, 314), vale decir, "*tenemos que amalgamar todo eso que... en el fondo tratar de rescatar lo posible de ellos*". (E02, 321-323). Nuevamente advertimos la expresión de rescate, aunque es el mismo entrevistado quien lo utiliza.

(b) Estrategias de agrupación en función de niveles de dominio.
La organización de los estudiantes en grupos es un aspecto crucial de la docencia según se desprende de las entrevistas. Los aspectos organizativos pueden ser una fuente que ayude al docente a estructurar una enseñanza más efectiva: "*entonces uno empieza a preguntarse dónde están las falencias y empieza a revisar la asistencia a aclases, a reorganizar un poco más el curso, con más apoyo y mayor incentivación a que ese apoyo que se les está dando lo tomen*" (E08, 375-380).

La agrupación puede ser una forma de enfrentar un proceso de reorganización de los grupos-curso. Encontramos dos estrategias: una, que busca dividir el grupo en secciones de niveles similares; la otra busca evitar este proceso de agrupamiento, favoreciendo la interacción de estudiantes con grados diversos de aprovechamiento académico.

(i) Agrupamiento como función de niveles similares de dominio.

Enfrentados a la realidad de estudiantes con diversos grados de conocimiento, algunos docentes plantean la necesidad de "*generar algún tipo de nivelación la forma de encauzar a los alumnos que tuvieran algún déficit mayor que otros, cierto, de forma de separarlos y tener otros tratamientos, hay varios proyectos*" (E02, 338-342). Dos conceptos se entrecruzan en esta expresión: el primero, denota una genuina preocupación por el avance de los estudiantes; el segundo, la idea pedagógica que los sujetos diferentes deben formar grupos diferentes, de manera de apoyarles de manera diferencial atendiendo a sus dificultades. Sin embargo, frente a la posibilidad "*de poder separarlos, o sea, lo ideal sería eso, pero no tenemos capacidad instalada de docentes como para que aquellos que van reprobando o tienen dificultades realmente grave, o sea, son problemas incluso de lectura, capaces de leer un documento e ir completando ese documento y derivarlos inmediatamente a un apoyo intenso*" (E02, 361-369).

La idea del tratamiento pedagógico por separado de acuerdo a grados variables de dominio en algunos casos se realiza de manera sistemática y "oficial" por así decirlo: "*en esta escuela estamos haciendo en matemáticas el primer día test y se separa el curso en dos partes, la mitad para abajo y la mitad para arriba y la mitad para abajo tiene la obligación de todos los días ir a un curso de nivelación para que se pongan al nivel de los demás, te fijas, en vez de bajar el nivel subimos de abajo para arriba y creo que eso, o sea, la palabra nivelación significa de abajo para arriba no de arriba para abajo [en ese aspecto la universidad hace nivelación] nosotros lo estamos haciendo, estamos haciendo un curso de nivelación en matemática*". (E03, 305-319). En este enfoque aparece con claridad la consideración del punto de partida de los estudiantes, sin caer en el desconocimiento que provoca la visión "estadística" de los estudiantes, asimilándolos a un promedio inexistente y fijando un punto de partida arbitrario.

(ii) Agrupamiento como función de colaboración.

El agrupamiento de los estudiantes como función de colaboración se produce al interior del aula, consistiendo en que los estudiantes más aprovechados en el desarrollo de una actividad específica entran a colaborar con sus compañeros que van a una velocidad menor. No corresponde, entonces, a una noción de alumno tutor o mentor que sistemáticamente acompaña a sus condiscípulos en el proceso de aprender. Específicamente, la experiencia es relatada en relación a la experimentación con módulos de autoinstrucción, en los cuales la velocidad o ritmo de aprendizaje es clave y los tiempos diferenciales juegan un rol fundamental para determinar el aprovechamiento (concepto muy fuerte en la teoría del dominio de J.B. Carroll y seguidores, que pone al tiempo como la variable clave del aprendizaje).

La experiencia más clara en este sentido es narrada por un docente que indica que "*intentamos hacer en un alto grado cuando estábamos tratando los módulos. Era más bien detectar la gente que se veía tenía un mejor rendimiento ¿ya? como para ir haciéndolo cabeza de los grupos ¿ya? para que ese mismo fuera polo de ayuda al resto de sus compañeros y no se produjera lo que normalmente se produce: que los malos se juntan con todos los malos y los buenos se juntan con los buenos ¿ya? Entonces una forma de que esta gente que tiene un mejor rendimiento sirviera de catalizador para el grupo o ayudara al resto de la gente pero a ese nivel no más, o sea, no de poner a los alumnos destacados ni cosa por el estilo, o sea, que la persona que está mejor preparara realmente ayude a sus compañeros ¿ya? de otra forma*". (E02, 505-523)

(c) Estrategia de experimentación didáctica.
Algunos docentes indican que han intentado innovar en los procesos didácticos, sea a través de la modalidad formal (Nota 3) existente en la universidad o por propia iniciativa. Los proyectos de innovación por vía formal disponen no sólo de una evaluación técnica sino que además se asigna una cantidad de recursos para financiarlos. Las variantes que asumen los proyectos son amplias. Por ejemplo, un docente señala que "*yo he desarrollado dos proyectos de mejoramiento docente en el sentido de que trabajar con módulos, o sea, entregarle material en el cual... guías... donde vayan desarrollando las guías con ayuda de ayudantes y profesores y posteriormente se les va evaluando esas guías como una forma de ir tratando de uniformar a toda la gente*" (E02, 351-359). Estos proyectos son evaluados y sistematizados anualmente.

Existen otras iniciativas no formalizadas, que transcurren por iniciativa de los docentes, pero carentes de las instancias de evaluación, sistematización y socialización. Tampoco queda claro en qué consiste propiamente la experimentación. En un caso, un docente señala "*experimento un par de años y veo si la cosa va por ese camino o no va por ese camino, el año pasado, antepasado hicimos ese sistema de módulos, por ejemplo, que era complemento a la clase de cátedra pero los concentramos todos el primer mes, primer mes intensivo exclusivamente los módulos, o sea, con algunas cuestiones básicas para nivelar en el fondo a toda la gente, por lo tanto es ese período eso hubo clase de cátedra propiamente tal, sino que una vez terminados los módulos todos eran evaluados y ahí se constituían, pero como el resto del año no había módulos la cosa se vino guarda abajo de nuevo*". (E02, 395-411). La variante que utilizó el mismo profesor fue de una entrega más distanciada en el tiempo. Señala que "*se probó el año pasado otro sistema, los módulos se iban haciendo semana a semana, o sea, se diseñaban como todos los programas normales, las carreras habituales, pero todos los días había, o sea un día a la semana había un módulo, que pasaba con eso, de que el tema de alguna manera se relajó porque la cosa aquí funciona muy bien los primeros meses, pero* *cuando empiezan las pruebas en otra asignatura de carrera, este alumno no está capacitado para hacer más de una actividad en el mismo tiempo y en el fondo se relajó"*. (E02, 411-425).

El uso de apoyo audiovisual es una práctica frecuente en la enseñanza universitaria. Sin embargo, de acuerdo a ciertas opiniones hay en los estudiantes una actitud de pasividad frente a la enseñanza, tal que "*constante, uno ha pasado por presentar transparencias y ha visto que las transparencias lo único que se dedican es a escribir lo que está en la transparencia, por lo tanto hemos colgado las transparencias, hemos vuelto a las clases clásicas, o sea, en el fondo uno va viendo si realmente ¿ya? de que forma ¿ya? puede dar mejor a, los alumnos*". (E02, 450-458). En el caso específico de este docente se advierte que la hipótesis inicial que sustentaba el uso de transparencias se derrumba cuando los estudiantes asumen que las proyecciones son sólo versiones sofisticadas de lo que habitualmente el profesor escribe en el pizarrón. Ante la situación, el docente no hace otra cosa sino volver a las prácticas habituales de uso del pizarrón. La clase frontal es la misma, por cierto, con o sin uso de transparencias o apoyos de este tipo.

(d) Estrategias para orientar el trabajo de los estudiantes.
Encontramos en las entrevistas una serie estrategias orientadas a mejorar, motivar, orientar el trabajo de los estudiantes.

(i) Motivación.

Algunos docentes tienen especial cuidado en que sus clases sean "entretenidas", interesantes, para los estudiantes de manera que éstos mantengan la atención. Un docente alude a su experiencia profesional tanto en el país como en el extranjero que le ayuda a la labor docente: "*yo creo que ese es lo que le hace que a mi me sea relativamente fácil enseñar y creo que mis alumnos se entretienen en mis clases y creo que aprenden*". (E03, 61-65).

También existen quienes utilizan las calificaciones como una herramienta de motivación, al menos para aquellos que tienen buenos resultados. Un profesor cuenta que "*yo la semana pasada repartí unas pruebas y a los que se sacaron más de 6 o 7, digamos, al entregar no solamente les llamaba por el nombre y les entregaba la prueba, les decía: 'fulano de tal un 7', para que todo el mundo vea que es posible sacarse un siete, esa es mi metodología*". (E03, 357-363). Por cierto, la práctica señalada es completamente conductista: no sólo provoca una respuesta en aquellos que nombra como estudiantes destacados, sino que también calla los nombres de otros. El silencio es una manera de descalificar, de desacreditar y de sancionar. Por cierto, la lógica del paradigma conductista lleva a pensar que los estudiantes calificados como de altos rendimientos por el profesor tenderán a mantener ese rango; en cambio, aquellos cuyos nombres fueron silenciados, probablemente seguirán en el silencio y en la oscuridad.

(ii) "Coaching"

Entendemos el "coaching" en el sentido original que propone Echeverría (1995), es decir, como el acompañamiento de un aprendiz por parte de alguien conocedor y experimentado en el tema, alguien que ya ha hecho el recorrido y está en condiciones de sugerir las vías más apropiadas para que el estudiante pueda recorrerlas sin el riesgo de desviarse (Nota 4). Por ejemplo, señala un profesor que él considera que puede ayudar a sus estudiantes más facilmente a "*hacer el camino, a descubrir, a que si yo los pusiera solos en una biblioteca, a decir: 'mire usted con un mes y veamos que pasó con estas primeras ideas liberales en América Latina', que seguramente sería un trabajo muy largo; en cambio, para mi enseñar sería acortarles ese trecho, indicarles donde están los textos, hacer la síntesis correspondiente*" (E04, 113-122). En un caso específico de enseñanza del derecho, el profesor indica que lo crucial es que los estudiantes aprendan a distinguir lo importante de lo sin importancia: "*la manera más obvia es diciendo lo que es importante y no importante. Muchas veces uno se estudia todos los contratos y va diciendo lo que es importante y no importante*" (E07, 66-70).

Asociado a lo anterior se encuentra el antiguo y siempre nuevo modelo de la interrogación ya sea socrática o simplemente interrogación como resultado de un proceso admirativo (en el sentido aristotélico). El incitar la duda, remover o sacudir estructuras mentales o axiológicas, abrir las fronteras, son todas acciones que colaboran a que el estudiante pueda abrirse a una nueva dinámica del saber: la pregunta, especialmente en un sistema que por doce o más años le ha enseñado que lo único correcto son las respuestas. "*A mi me ha dado mucho resultado en las clases cuando ellos me ven a mi dudando de muchas cosas; el profesor se humaniza de alguna manera y el alumno se entusiasma más. Más que tener al profesor que habla y habla y habla y sabe, sabe todo, le preguntan, respuesta, le preguntan, respuesta, seguro que la pregunta es muy fácil, él sabe todo entonces como hacer, yo entiendo que hay disciplinas en las que se puede y otras en las que no se puede, pero cómo hacerlo para que el profesor investigue en su propio conocimiento, eso es muy interesante*". (E04, 636-650). La doble dimensionalidad que presenta la cita es destacable: el profesor entrevistado no señala que la pregunta es algo que debe existir en el estudiante sino que también en el que enseña. El acto de preguntar, como siempre ha sido, es el acto fundamental del inteligir de la realidad. No se puede ser maestro ni estudiante sin la capacidad de preguntar. Al hacerlo ampliarán sus horizontes y entonces "*ya pueden empezar a relacionar y a ser gozosos con el conocimiento*". (E04, 260-261).

El acto de preguntar, sin embargo, no es mecánico ni se puede decidir de la noche a la mañana, sino que "*tiene que haber una sensibilidad hacia los estudiantes*". (E05, 56-57), una capacidad instalada y activa de escucha "ontológica" de la voz de los estudiantes (Echeverría, 1995).

(e) Estrategias de búsqueda de significación.
Algunos docentes han encontrado que una forma eficiente y eficaz de enfrentar la enseñanza proviene de la conexión que ellos puedan hacer –y llevar a los estudiantes a hacerlo- entre lo que enseñan y las preocupaciones y realidades de la vida diaria o de la profesión para la que se están formando los estudiantes. Como señala una docente, "*estoy todo el tiempo mirando el diario, buscando... entonces yo uso eso, uso las revistas del Sernam como fuentes de información*" (E05, 266-269) a partir de las cuales desarrolla los problemas de trabajo en su disciplina, de donde los estudiantes puedan asociarla con situaciones de la vida

diaria.

Una alternativa a lo anterior la plantea la misma entrevistada cuando expresa que la significación puede encontrarse también con respecto a otras asignaturas que estén cursando. Cita el caso en que los estudiantes tuvieron "*una tarea en ecología y yo justo tocó que los alumnos estaban haciendo estadística y ecología al mismo tiempo y entonces la resolvimos juntos*". Por cierto, esta posibilidad enriquece y fortalece la docencia, si bien no siempre es posible encontrar las oportunidades para establecer tales vínculos, lo que puede constituir todo un desafío para una docencia que busca superar los moldes tradicionales de la rutina y la dependencia.

En las expresiones anteriores encontramos una propuesta de aprendizaje funcional con significado. La profesora se preocupa es establecer relaciones que asocien lo aprendido en la propia asignatura (i) con eventos de la vida diaria, de la cotidianeidad, tanto de la profesión como de la vida ciudadana, y (ii) con otras disciplinas en las que se encuentran involucrados los estudiantes.

El proceso implicado en esta modalidad estratégica es descrito como "*trabajos prácticos que buscan que el alumno llegue al concepto, y ha sido muy rico cómo ellos llegan a él sin que uno haya dicho nada y uno después los formaliza. No sé, eso es una experiencia realmente exquisita. Entonces, sí, ¿cómo producir esa magia? ¿cómo ver a los alumnos descubriendo?*" (E05, 349-356).

(f) Establecimiento de contratos.
Una alternativa estratégica señalada por un docente es establecer relaciones de contrato o compromiso entre los estudiantes y el profesor, entendiendo que con ello los alumnos estarán más incentivados para el aprendizaje. Así, según un profesor se "*puede hacer un compromiso con ellos, de tal manera que se incentiven, se preocupen. Yo creo que rinde más, yo creo que es posible*" (E08, 423-426).

La importancia de este aporte radica de manera principal en el desplazamiento del control. Si el aula es considerada como un espacio político, en ella pueden identificarse lugares donde se concentra el poder, formas de negociación y de mantenimiento de la estructura relacional (Ball, 1987). En este caso, el docente propone la transferencia de una cuota de poder hacia los estudiantes, con lo que no sólo promueve la autonomía intelectual de los mismos sino que comportamientos ciudadanos de primera importancia para la vida profesional futura.

(g) Estrategia de aprender haciendo.
El aprender por la acción tiene claras implicancias en la docencia de algunas asignaturas, específicamente aquellas relacionadas con la tecnología o procedimientos. Al respecto, un docente señala que cuando "*un auditorio escucha una clase lectiva, de eso más o menos capta el 20%, el 25%; si lo que él ve, las cosas que explica el profesor mediante gráficos, fotografías, ahí adquiere un 40%; y si él trabaja en algo y se ejercita en algo ¿no es cierto?, hace un trabajo, aprende un 80%*" (E03, 86-92). Más allá de lo exactos que puedan ser estos porcentajes, la expresión trasunta una convicción que siempre ha estado presente en las teorías del aprendizaje y la enseñanza y que en la actualidad ha recibido una fundamentación sustantiva de parte de la psicología sociocognitiva constructivista.

Hay un caso específico de una docente que enseña una asignatura del campo tecnológico, quien señala que "*en cambio, con nosotros el alumno tiene que trabajar uno por uno, y es evaluado en el momento y cada cosa que haga tiene que ser evaluada; no se puede ir. No le podemos hacer nosotros una demostración. Es eminentemente práctica mi asignatura y es lo que tiene que hacer después en su desarrollo profesional*" (E12, 51-58).

Otro docente, ahora del campo de las ciencias naturales, plantea una posición similar. Dice que las "*clases de física deberían ser todas, prácticamente todas salvo algunos cursos de física más teórica, digamos, deberían ser con experimentos demostrativos durante la clase. [...] en los laboratorios los alumnos manipulan las cosas. Eso es otra cosa y, naturalmente que también es necesario, sobre todo en carreras de corte técnico como ingeniería*" (E14, 276-289). Aparece claro el sesgo experimental del docente, pero lo más señalado es la insistencia acerca de la manipulación experimental que deben hacer los estudiantes durante el laboratorio, así como la claridad con que plantea que ello es especialmente necesario en las carreras más asociadas con las tecnologías duras.

4.3. Fuentes de dificultad.

El desarrollo de la docencia, cualquiera sea la modalidad estratégica que se asuma, no está exento de dificultades. Más aún, podría decirse que cada modalidad o enfoque estratégico tiene sus propias dificultades intrínsecas tanto como originadas en las condiciones contextuales y específicas en que se realiza la tarea docente.

Una de las restricciones fuertes que se encuentran, especialmente en los cursos iniciales, tiene que ver con el número de alumnos. Como lo señala un docente, "*uno tiene que ir muchas veces ajustando eso en función de la cantidad de alumnos que tiene, pero no siempre se puede, por ejemplo, pedir trabajos individuales a noventa alumnos porque uno es incapaz de... y es incapaz de presentarlos tampoco en clases*" (E08, 455-461).

El concepto de "gran número de alumnos" es relativo, dependiendo de cual es la estrategia docente privilegiada. Un trabajo docente orientado a laboratorio o taller (independiente que sean duros o blandos) va a demandar grupos más pequeños. Pero, si la estrategia es fundamentalmente la conferencia, entonces el problema no radica en la cantidad de estudiantes, excepto en el caso de no disponer de salones de mayor capacidad o de sistemas de amplificación de la voz. También puede suceder que el argumento del exceso de alumnos sea para evadir cualquier iniciativa orientada al cambio de los paradigmas docentes habituales.

Otra restricción importante se asocia a la escasez "*de tiempo. También es cierto que hay poco espacio* (temporal) *para hacer innovaciones*". (E11, 415-416). El tiempo disponible (tantas horas de docencia por un número determinado de semanas) no sólo actúa como regulador de las secuencias instruccionales sino que aparece como un regulador de los contenidos a enseñar, los que serán seleccionados (por el docente o por el departamento académico) en función de dicha disponibilidad.

4.4. A modo de síntesis.

Los diversos elementos recogidos y estructurados en la comunicación anterior pueden sintetizarse en los siguientes acápites.

- Hay conciencia de la falta de preparación didáctica de los docentes de la universidad;
- La enseñanza se concibe como transmisión, orientada al cambio conductual; también hay docentes que conciben un rol de guía, estimulación y modelamiento; sin embargo, siempre se mantiene la asimetría de la relación profesor-estudiante;
- Entre las estrategias utilizadas se destacan las de
-
    - nivelación de grupos;
    - agrupación en función de niveles de dominio;
    - experimentación didáctica;
    - orientación tanto por motivación como por "coaching";
    - búsqueda de significación;
    - establecimiento de contratos;
    - aprender haciendo;
- Los docentes manifiestan que el excesivo número de alumnos así como las restricciones en el tiempo de la docencia constituyen fuentes de dificultad para el ejercicio de su trabajo.

5. Conocimientos previos.

Este tema es una preocupación abierta en la pedagogía actual. En el campo universitario constituye un problema de la cotidianeidad de la tarea docente, en la cual los profesores se enfrentan a estudiantes con antecedentes diferenciales que obedecen a diversas fuentes. La complejidad de la temática queda en evidencia en la Ilustración 9. La mayor variabilidad que presenta la ilustración respecto de anteriores indica un grado menor de consenso entre los docengtes respecto del punto.

En este apartado se intenta responder preguntas relativas a (i) la importancia que le asignan los docentes a la evaluación de conocimientos previos, (ii) los propósitos que podría tener dicha evaluación, (iii) la forma en que visualizan la problemática según características de los estudiantes, y (iv) los efectos que se hacen sentir sobre la docencia.

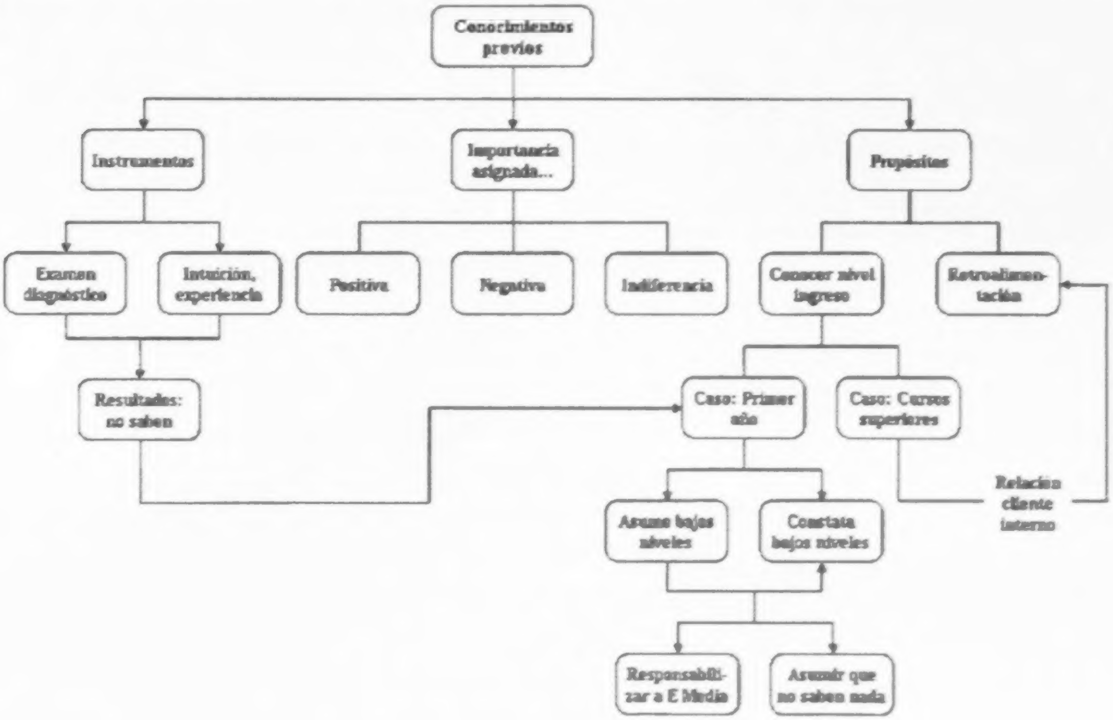

Conocimientos previos
Instrumentos
Importancia asignada...
Propósitos
Examen diagnóstico
Intuición, experiencia
Positiva
Negativa
Indiferencia
Conocer nivel ingreso
Retroalimentación
Resultados: no saben
Caso: Primer año
Caso: Cursos superiores
Relación cliente interno
Asume bajos niveles
Constata bajos niveles
Responsabilizar a E Media
Asumir que no saben nada

Ilustración 9. Conocimientos previos

5.1. Importancia de su evaluación.

Se consultó a los docentes si consideraban importante evaluar los conocimientos previos de los estudiantes. Una respuesta típica es "*Sí, yo creo que sí, para saber sobre que base uno puede partir, claro porque muchas veces recibe cursos que son muy heterogéneos, que vienen con este nivel, con este otro nivel, con esta otra formación, todos se juntan*" (E12, 448-453).

La importancia asignada es alta, tanto así que en una unidad académica adoptaron internamente como norma el evaluar diagnósticamente a los estudiantes. Un académico señala que "*Sí, de hecho cuando en el departamento de administración nos pusimos como norma que (parte se cumplió y en otras no) que cada profesor hiciera una prueba de diagnóstico en los 15 primeros minutos al inicio del semestre para poder saber, en términos vulgares, el lugar que ocupa, como estaba, como partía.*" (E01, 331-338). Entonces, la temática de conocer el nivel de conocimientos con que los estudiantes ingresan a una determinada asignatura se convierte en una norma importante para los docentes.

Existe también una posición que considera que la evaluación de los conocimientos previos es un ejercicio ocioso, no porque no sea necesario conocer el nivel de ingreso de los estudiantes, sino porque consideran que poseen hipótesis lo suficientemente fuertes como para poder obviar el proceso. A diferencia del "*profesor de primer año que ya lo obtuvo, quizás él requiera un diagnóstico previo, pero nosotros prácticamente tenemos alguna clases de diagnóstico y realmente sabemos cual es el material que nos vamos a encontrar*". (E11, 301-306). Otro, de manera similar, expresa que "*para el tercer nivel de mi curso no, primeros años creo que sí*" (E08, 338-339).

Por cierto, también hay quien no concibe la evaluación de los conocimientos previos o bien no tiene interés en ello o, incluso, no tiene idea del sentido de la misma. Señala uno de ellos: "*No sé para qué, la verdad es que no sé. Hay un punto... bueno la verdad es que no sé para qué, pero lo hago, lo hago, pero la verdad es que no sé para qué lo hago. Lo hago al iniciar cada curso, en la primera sesión yo les hago un test. ¡Ah! se horrorizan por, dicen, es un test sin nota.*" (E02, 260-267). La expresión anotada también hace referencia a la actitud de los estudiantes, quienes, según el docente, se opondrían a un examen que no llevara calificación; esto estaría en contradicción con la tradición, con una modalidad formativa durante más de doce años, en que cualquier prueba o examen está relacionada con una calificación.

En todo caso, no queda claro el sentido de evaluar los conocimientos previos, ni el uso que se da a la información. De hecho puede suceder (i) que se utilicen los resultados para establecer puntos de partida próximos o ligados a los reales conocimientos de los estudiantes; (ii) que se los mida, pero no se los utilice para ajustar la enseñanza sino sólo para establecer el grado de la ignorancia; (iii) que se los mida pero no se los utilice con ningún propósito.

5.2. Propósitos de la evaluación de conocimientos previos.

Esta tiene en la visión de los docentes al menos dos propósitos: conocer el estado de ingreso de los estudiantes y retroalimentar a los docentes del período anterior. Como lo expresa claramente uno de ellos, "*eso tenía dos objetivos. Uno, primero es saber eso, y el otro es poder retroalimentar a los profesores que venían de antes. Decirles: 'mire profesor fíjese que el curso que usted dejó pasar, o sus alumnos no saben tal cosa, por así decirlo; no merece salir de la carrera, no está capacitado para salir porque esto que es fundamental no lo sabe'*" (E01, 338-347).

El primer propósito se plantea en al menos dos niveles: al nivel del primer año y en los cursos superiores. En los estudiantes de primer año se examina en párrafos posteriores. Tratándose de estudiantes de cursos superiores se advierte una hipótesis subyacente y que implica que los que ingresan a una asignatura ya han pasado por otras que son requisitos de la presente y las han aprobado, así se infiere que han sido nivelados y poseen las condiciones para aprobar.

El segundo propósito es lo que se denomina la relación del "cliente interno", concepto que es utilizado por los docentes, donde cada fase del proceso debe ser realizada con un nivel de calidad tal que asegure que al final el producto cumplirá los estándares requeridos para su ingreso al mercado. Análogamente, cada profesor debe procurar que sus estudiantes salgan tan preparados que puedan cursar la asignatura siguiente de manera eficiente, sin lagunas de conocimiento, para que al final del proceso la universidad pueda acreditar ante la sociedad a una persona para el ejercicio de una profesión determinada.

Un docente expresa que la evaluación de conocimientos previos "*de hecho no un diagnostico pedagógico porque yo trato de ligar mis asignaturas con los aprendizajes de las asignaturas anteriores; entonces no hago un diagnostico; quizás debería tomar una prueba de conocimiento que son con los cuales está directamente relacionada; quizás pudiera preguntar por los conceptos que pudieran estar presente; quizás muchas veces haya diferencias; quizás muchas veces no respondan; quizás sirviera para ubicarlos en la sabiduría o ignorancia de las materia previas y la verdad que seria útil.*" (E09, 202-215). A pesar de la utilidad que se declara, de hecho no se hace una evaluación de conocimientos previos, según se desprende de las propias palabras del entrevistado.

En otra entrevista el profesor señala: "*porque yo creo que una asignatura que es continuación de otras, asignaturas que se dictan por profesores que de aquí mismo, -nosotros conversamos mucho-, entonces yo sé exactamente cual es el nivel, o exactamente no, con un nivel de exactitud cuál es el nivel de cada curso. ¿Por qué? Porque ellos provienen de un determinado curso de biología que se da el primer año con un profesor que está en la oficina de más allá, entonces él me dice 'mira estos cabros (Nota 5) vienen aquí, tienen tales problemas'. Más o menos sabemos con qué material nos vamos a encontrar y generalmente es acertado*" (E11, 273-287). En este caso la información acerca del estado de los estudiantes no proviene de los resultados de un examen que se les haya aplicado sino que del mismo docente que les enseñó en el período anterior. Sin embargo, esta relación es posible tan sólo cuando se trata de docentes que pertenecen a la misma unidad académica. Ahora bien, dado que la organización de las escuelas profesionales es de carácter matricial, puede suceder perfectamente que dos profesores que enseñan materias relacionadas no tengan ninguna relación entre ellos. Con ello, entonces, carecerán de toda articulación, llevándose a cabo una enseñaza basada en supuestos gratuitos.

En una línea similar a la anterior, otro profesor dice que "*Hay veces que sí y hay veces que no, hay cursos que son secuenciales, de modo que si aprobó la física uno, por ejemplo, tomó la física dos y si yo voy a dictar la física dos tengo una idea más o menos clara de lo mínimo que saben esos alumnos, entonces creo que ahí casi puede ser innecesario*" (E14, 206-213). En este caso, no obstante, la información no proviene de los resultados de los estudiantes en un examen, ni tampoco se origina en el profesor que enseñó el curso anterior. Existe la tercera posibilidad que se relaciona con la estructura del currículo: el examen del programa de Física I le permite al profesor de Física II tener "*una idea más o menos clara de lo mínimo*" que han de saber los estudiantes si es que han aprobado Física I.

5.3. El caso de los alumnos de primer año.

Los alumnos de primer año representan una incógnita y un desafío para los docentes universitarios. La incógnita tiene que ver con cuánto saben o no saben, cuánto son capaces o incapaces de hacer, qué actitudes tendrán hacia el estudio, su capacidad de autonomía y administración de la propia libertad en un ambiente radicalmente diferente al liceo. El desafío proviene de tener que enseñar un programa, generalmente de ciencia básica de las profesiones, cuyos requisitos de entrada exceden habitualmente a los estándares de los estudiantes. Esta actitud (ya sea institucional o personal) implica no prestar atención a los niveles de entrada de los estudiantes y pone la atención más sobre la enseñanza más que sobre el aprendizaje.

Ante lo inevitable de la situación, una primera mirada es la de responsabilizar a la enseñanza media de los bajos niveles de sus egresados. La evidencia a que pueden aludir los docentes no sólo proviene de su propia experiencia personal enseñando a estudiantes de primer año o de los exámenes de diagnóstico, sino que hay evidencia suficiente a partir de los exámenes de admisión acerca del bajo nivel de dominio que tienen los estudiantes egresados de la enseñanza media sobre las materias fundamentales. Como lo señala un docente, de seguir las cosas como están en la "*enseñanza media, pues al final la enseñanza media no va a aprender nada pensando que una vez entrado a la universidad ahí que me enseñen matemática, física y química, o en inglés, qué sé yo*" (E04, 411-415). No es posible inferir de aquí una responsabilización de la institucionalidad escolar de la enseñanza media sino que una referencia a que algo no ha funcionado apropiadamente durante el proceso de formación, sin llegar a identificar cuáles han sido los elementos y cómo se han combinado para producir tal efecto.

Todo lo anterior tiene un efecto sobre la universidad, llegando al punto que un profesor señala que "*yo podría decir simplemente si corresponde que la universidad tape hoyos que vienen de abajo y en esa primera parte por cierto que también hay un debate; algunos académicos te dicen, bueno nosotros no podemos hacer el quinto medio o el sexto medio, porque les estamos facilitando cada vez más las cosas a la enseñanza media*" (E04, 403-417), lo que vendría a desvirtuar todo el planteamiento de lo que significa la educación superior. Con todo, la visión acerca del futuro no es positiva, máxime si se considera la oferta creciente de vacantes en la educación superior: "*yo tengo la impresión que por una cuestión objetiva como la gran existencia de universidades eso se va a transformar en un hecho, las universidades van a tener que empezar a suplir seriamente los problemas de falencia, de conocimientos que llevan los alumnos*" (E04, 417-424). Pero no es sólo una crítica a la enseñanza media, sino que también constituye una clara negación de los estudiantes y sus personales características, una desvalorización de sus conocimientos previos, un desvirtuamiento del esfuerzo de doce años de estudio, y su nivelación por el rasero de la "igualación a cero".

Ante esta situación se genera una fuerte presión sobre los docentes, quienes se ven llevados a reaccionar de alguna forma. La imagen más común que se tiene es que los estudiantes saben poco menos que nada y que, en tal caso, lo mejor es hacer como que nada saben y comenzar a enseñar desde cero, desde el inicio, como si se tratara de materias completamente nuevas.

5.4. Alumnos de cursos superiores.

La situación respecto de los alumnos de cursos superiores no se presenta mejor. De hecho, las opiniones de los docentes en relación a la preparación de los estudiantes que reciben siempre son más negativas que positivas. De acuerdo a un docente que enseña en cursos superiores, "*en cuarto año, la verdad es que no necesito hacer un diagnóstico porque yo tengo más o menos de cuanto llegan sabiendo de filosofía, porque básicamente de cero, o sea sabiendo cero de filosofía y teniendo cero entrenamiento para enfrentarse con cuestiones de una relativa sofisticación filosófica, la verdad es que cualquier diagnóstico que yo haría me daría sino cero, cero punto uno y no es mucha la diferencia.*" (E07, 411-421).

Lo notable de la referencia anterior es que en el fondo es la misma crítica que hacen los docentes del primer año sobre la enseñanza media. Ahora se trata de alumnos de cuarto año, de los cuales puede esperarse que hayan logrado las competencias necesarias para desempeñarse en ese nivel: mal que mal, han sido promovidos durante tres años. Pero sucede que luego de tres años de arduo entrenamiento en las aulas universitarias, todavía permanece la misma queja. Esto no parece consistente. Ya no es posible responsabilizar a la enseñanza media. Restan solamente la propia institucionalidad y el estudiante como sujeto individual.

5.5. Instrumentación.

Asumiendo que es relevante evaluar los conocimientos previos de los estudiantes, la pregunta es ¿con qué medir? La respuesta más habitual sería con un examen o prueba escrita. De hecho se habla de una "*prueba de diagnóstico en los primeros 15 minutos al inicio del semestre*" (E01,334-336), así como también de "*un test o una prueba*" (E06, 239). En breve, la confianza casi ilimitada, ya sea por falta de conocimiento o por falta de imaginación, en los instrumentos de carácter escrito, en que el estudiante debe producir respuestas más que preguntas, debe converger antes que divergir, debe ceñirse a los saberes consagrados antes que desafiarlos. La referencia a 15 minutos que hace un docente ayuda a encuadrar la visión que se tiene de la naturaleza de una evaluación de conocimientos previos.

Pero también hay quienes confían en la intuición o su propia experiencia. Dice un profesor: "*entonces, la verdad, no necesito hacer cada año un diagnóstico porque ya más o menos los ubico*" (E07, 407-409). Otro indica que "*el hecho que un alumno ha hecho una asignatura u otra hace la diferencia y eso se percibe inmediatamente*" (E10, 217-219). También se dice que a "*veces ese diagnóstico no es necesario que sea a través de un test o una prueba, yo creo que puede ser también a través de una percepción que se va haciendo a lo largo del curso y uno tiene que ir*" (E06, 238-242). El "impresionismo" con que reaccionan algunos docentes queda claramente expresado en "*tengo la impresión de que por lo menos en filosofía no varían mucho, digamos, y habitualmente siempre es así*" (E07, 425-428). Esa expresión de "tengo la impresión" permite visualizar de manera definitiva el alto grado de subjetividad que impregna estos juicios, aunque no puede decirse que es infundada.

Por otra parte, no deja de llamar la atención la confianza que estos docentes ponen en sus propias percepciones en lugar de acudir a instrumentos u otra estrategia de evaluación más confiables. Sin embargo, puede suceder que los instrumentos elaborados por ellos mismos y luego administrados a los estudiantes produzcan resultados completamente previsibles. Pero también debe considerarse que los instrumentos estarían midiendo lo que el profesor quiere que los estudiantes sepan y no lo que éstos saben realmente. Esto implica desafíos nuevos para los docentes en términos de diseñar estrategias de evaluación que les permitan no sólo conocer sino hacerse cargo efectivamente de lo que los estudiantes conocen y no conocen.

5.6. Tipos de déficit en los alumnos.

Los déficit detectados en los estudiantes pueden agruparse en conocimientos y destrezas intelectuales. Volviendo al tema de los alumnos de primer año, se hacía referencia a que "*al final la enseñanza media no van a aprender nada pensando que una vez entrado a la universidad, ahí que me enseñen matemática, física y química*" (E04, 411-415). Similarmente, una docente plantea que cuando los estudiantes llegan hasta "*nosotros ya deberían tener un background de conocimientos cuando llegan acá, como le digo todas esas asignaturas, tienen que saber anatomía, biología, matemáticas, estadística, fisiología, se supone que ya las hicieron ya, debería ser en la parte básica*" (E12, 485-491). Aquí hallamos dos puntos de vista: uno de negación o sub-valoración de lo que los estudiantes han vivido durante la enseñanza media, y otro de lo que esperan que debería haber ocurrido en ese ciclo de la formación. Sin embargo, ninguna de estas posiciones se acerca verdaderamente a lo que el estudiante es, con toda su realidad, sus experiencias tanto dentro como fuera de las aulas escolares, con una familia en la que muchas veces él o ella es el primer profesional.

En segundo lugar encontramos la observación referida a procesos y destrezas intelectuales. Señala un docente que cuando comienza "*a trabajar con ellos se da cuenta que hay muchos problemas de método. Claro, uno no hace un diagnóstico pedagógico antes sino en el momento de recibir la evaluación*" (E08, 360-365). Al parecer, encontramos una docencia que se inicia en el vacío; pero eso es tiempo y esfuerzo perdido, no sólo del docente sino de los estudiantes. Al respecto es importante considerar lo que plantea Ausubel en términos de lo relevante que es para la docencia el conocimiento de lo que el alumno es y sabe, y actuar en consecuencia con ello.

5.7. Efectos sobre la docencia.

Ciertamente el tener información o no acerca de los conocimientos previos de los estudiantes de acuerdo al nivel en que se encuentran, así como el tipo de instrumental que se utiliza, provoca consecuencias en el plano docente. Entre éstas las más notables son las de disponer un recurso para retroalimentar a los colegas, el ajuste de la propia enseñanza a las características de los estudiantes, la reformulación o re-ajuste del currículo.

El efecto de retroalimentación ya ha sido mencionado bajo el concepto de "cliente interno", con todas las restricciones que pueda tener una analogía como la hecha: "*se trata de retroalimentar a los profesores que venían de antes*" (E01, 340-341). Por cierto, es bueno tener con qué retroalimentar a los otros docentes, pero más importante aún es contar con información para diseñar el trabajo que se realizará conjuntamente con los estudiantes.

En segundo término, los docentes pueden mejorar su enseñanza ajustando los tópicos (contenido, secuencia, ritmo) a las características de los diferentes grupos según estén "*preparados, entonces eso a lo mejor me ubicó para tratar de empezar la explicación de la materia*" (E02, 286-288), o como expresa otro profesor, "*me interesaría si es que hay un mecanismo para hacer un diagnóstico que me sea útil para cambiar la forma cómo enseño de tal forma que el proceso de aprendizaje sea más productivo*" (E15, 337-342). Por cierto, también la negativa a considerar los conocimientos previos o la confianza en la propia intuición son recursos con los cuales se decide no sólo el tipo de enseñanza sino que también el contenido de la misma. Este docente muestra un claro interés en modificar lo que hace, lo que es valioso y una puerta abierta para comenzar la innovación.

5.8. A modo de síntesis.

Los principales conceptos referidos a los conocimientos previos encontrados en las opiniones y expresiones de los docentes universitarios se pueden resumir en los siguientes:

- Hay opiniones diversas en torno a la importancia de evaluar los conocimientos previos de los estudiantes: desde quienes dicen hacerlo y utilizar los resultados para determinar un punto de partida, hasta quienes declaran hacerlo pero no tener idea del para qué de esa evaluación;
- No queda claro si la evaluación de conocimientos previos tiene algún impacto sobre la enseñanza (selección de contenidos, secuenciación de la instrucción, ritmo de trabajo, tipo y diversidad de la ejercitación, vinculación con la realidad y la profesión, etc);
- Hay una doble visión de los alumnos de primer año; por una parte, se desconocen sus características, sus conocimientos previos, sus historias; en otras palabras, se les hace invisibles; por la otra, al partir de la base de que nada saben, se

desvaloriza completamente lo que pudieran saber;
- En el caso de los alumnos de cursos superiores se plantean también problemas de conocimientos previos insuficientes, pero ahora el referente no es la enseñanza media sino que la propia universidad;
- La evaluación, cuando se hace, se hace de manera muy poco técnica, más bien intuitiva.

6. Actitudes Docentes

El análisis de esta sección está orientado por preguntas tales como cuáles son las actitudes consideradas "apropiadas" por los docentes universitarios, y cuáles serían las no apropiadas a la función, cuáles son las actitudes que se observan en la práctica, cómo se plantean los docentes hacia los estudiantes así como en referencia a la propia tarea de enseñar. La presenta gráficamente la estructura encontrada.

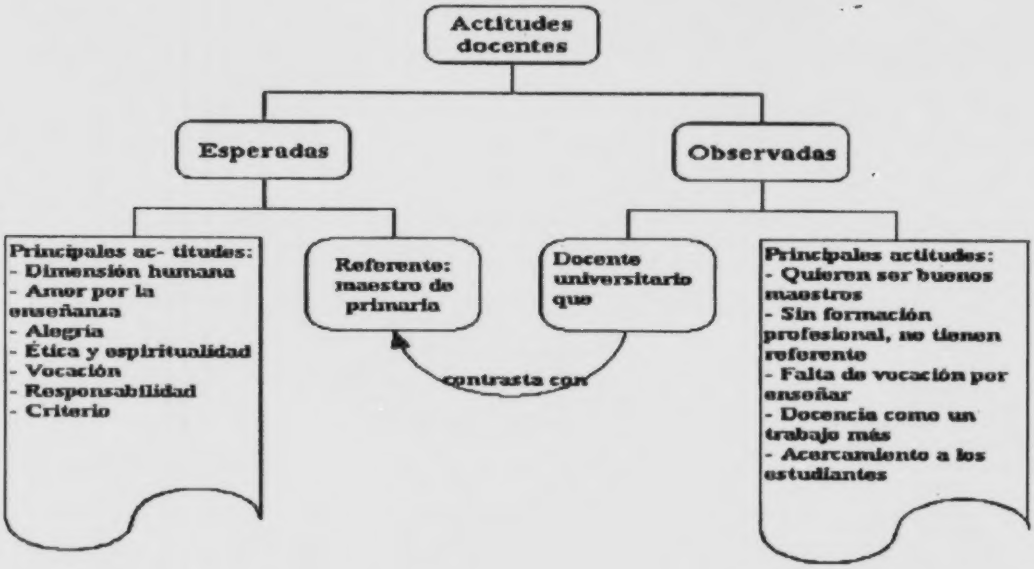

Actitudes docentes
Esperadas
Observadas
Principales ac- titudes:
- Dimensión humana
- Amor por la enseñanza
- Alegría
- Ética y espiritualidad
- Vocación
- Responsabilidad
- Criterio
Referente: maestro de primaria
Docente universitario que
contrasta con
Principales actitudes:
- Quieren ser buenos maestros
- Sin formación profesional, no tienen referente
- Falta de vocación por enseñar
- Docencia como un trabajo más
- Acercamiento a los estudiantes

Ilustración 10. Actitudes docentes, esperadas y observadas

6.1. Las actitudes esperadas en los docentes.

(a) Un referente.
A través del análisis de las respuestas es posible perfilar las actitudes esperadas de un docente desde su propia visión como practicantes de una profesión que no es la propia. Existen variadas respuestas, de las cuales la más destacada es aquella que colabora a la construcción de un referente: cuál es el espejo contra el cual se miran los docentes en su tarea de tales. En esta relación especular es donde luego podrán contrastarse las contradicciones vitales y conceptuales que se vayan dando en el ejercicio de la docencia, en el crecimiento y desarrollo de la función. En otras palabras, el efecto de modelaje, no sólo para los docentes sino modelaje de una actitud valiosa de vida para los mismos estudiantes.

Un segundo punto es que se trata de la visión de lo docente, no de la función académica completa, de donde quedan en la penumbra los componentes relacionados con la investigación científica y tecnológica, la creación artística, la difusión, la asistencia técnica, la gestión.

Encontramos en las entrevistas una larga exposición acerca de un referente de la docencia, que es el profesor primario o básico. Un largo texto expresa el sentimiento del entrevistado: "*yo recuerdo mi profesor básico, por ejemplo, en el colegio público en el que revisaba las manos, las orejas, las uñas, los pies, se preocupaba como era la cosa en la casa, cómo estudiaba el niño y si podía estudiar o no podía estudiar, si comía o no comía, todas esas cosas eran parte de la vida del profesor y eso tal vez en la enseñanza media pasaba algo similar. Yo creo que en la enseñanza en general de hace algunos años el profesor era también así y uno recuerda, que sé yo, por así decirlo a unos grandes personajes de la educación como Juan Gómez Millas, por ejemplo, u otros, y en distintas escuelas personas que prácticamente hacían clases por el orgullo de ser profesor, por el amor a que otros aprendan lo que ellos sabían*". (E01,

La anterior es una imagen retocada por el tiempo, endulzada por la añoranza de una niñez que se ve a la distancia como una época de felicidad, comparada con las complejidades y dificultades de la vida adulta. También es una época en que la educación pública gozaba de un prestigio que no había sido arrinconada por la arremetida de la educación privada, relegada al rincón de la ineficiencia e inoperancia. Los profesores a los ojos de los niños no son lo mismo que a los ojos adultos. Hay aspectos importantes que han de rescatarse de esta memoria.

(b) Actitudes esperadas.
Las actitudes esperadas o ideales de acuerdo a los docentes de la universidad son:

- La primera es que los recuerdos no hacen referencia a lo mucho que sabían sus profesores, sino a la preocupación personal del docente por aspectos "no académicos", o no profesionales tales como el aseo personal, la situación familiar, las condiciones de estudio.
- Una segunda dimensión son las actitudes, que en este recordatorio destacan en los auténticos profesores: orgullo por la profesión, y amor por la enseñanza. Con relación a esto, el mismo entrevistado agrega, quizás a manera de crítica o quizás como resignada comprobación de la realidad que "*hoy día de alguna forma también se ha transformado en una forma de ganarse la vida y no por amor. Ha ido cambiando...*" (E01, 535-537), perdiéndose aquello de antes cuando "*los que hacían las clases en enseñanza básica y enseñanza media eran como apóstoles de su disciplina*" (E01, 510-512).
- Se suma la alegría como un componente importantísimo de la profesión docente: "*realmente hacer un trabajo con alegría, no solamente porque hay que hacer clases*" (E01, 547-548).
- También es importante la dimensión ética y espiritual, que deben trasuntar la vida y acción del docente. "*Primero, el docente tiene que tener valores éticos altos y una espiritualidad que lo motive*" (E03, 220-222).
- La vocación por la enseñanza es una condición muy importante, tanto que para un entrevistado es casi como un eje vital. En efecto, "*yo creo que el buen profesor lo hace la vocación, lo hace el estar feliz de enseñar, el considerar 'tengo el mejor trabajo del mundo*" (E03, 378-382).
- Hacerse responsable de sus estudiantes, de su progreso y éxito se presenta como una característica adicional deseable del docente universitario. Según un docente, existe "*una responsabilidad de cada profesor en el nivel donde está de poder superar algo, a lo mejor, las deficiencias que pudieran traer los alumnos*" (E06, 95-98).
- Una característica adicional viene dada por el criterio, esa forma moderada y ponderada de enfrentar las cosas, los problemas y las dificultades especialmente: "*una persona que tiene un criterio flexible, elástico, y sabe adaptarse*" (E08, 268-270).

6.2. Las actitudes observadas en los docentes.

Conceptualmente, las "actitudes observadas" corresponden a las referencias que hacen los entrevistados a las prácticas concretas, tanto propia como ajenas, del ejercicio docente y cómo interpretan ellos esas prácticas en términos de actitudes. Consiguientemente, el término "observadas" no implica que se hayan hecho observaciones orientadas específicamente a detectarlas, sino que son observadas para los entrevistados, quienes dan cuenta de ellas.

La primera y fuerte constatación es la contraposición que se hace del modelo ideal del docente (representado por aquel profesor de la niñez...) con lo que sucede hoy en las aulas. La comparación más violenta es la que se hace entre el sentido apostólico de la vocación docente y el vulgar ejercicio de una profesión para ganarse la vida. Se señalaba que la docencia se ha transformado en nuestros días en "*una forma de ganarse la vida y no por amor*" (E01, 537-537), "*lo que le interesa es ganar plata, no más*" (E03, 223-224). Por ello se plantea que el profesor debe tener "*una fuerte motivación de enseñar y no un profesor que se esté ganando un sueldo; o sea, yo creo que eso hace una gran diferencia: el profesor que se está ganando el sueldo, además, los alumnos lo van a notar*" (E03, 373-378).

Voluntad de ser entendida como "*lo que ellos consideran buenos profesores para ellos quieren serlo para sus estudiantes*" (E01, 422-424). En opinión de este entrevistado, no se trata que los docentes eventualmente sean defectuosos o inhábiles porque quieren serlo, sino porque no tienen preparación, aunque puede haber algo de "*inercia, hablando en término general de los docentes: muchos docentes hacen clases como ellos recibieron sus clases y las han hecho durante un montón de tiempo*" (E11, 411-415). El racional implícito en esta expresión es que de hecho, todo profesor quiere ser un buen profesor y para ello, careciendo de formación profesional en el área, entonces busca reproducir aquellos modelos que ha conocido. Esto acarrea que se conserve con mucha fuerza el modelo tradicional, "*el esquema autoritario, por así decirlo de alguna manera, anterior está vigente. El profesor sabe... eh... y el alumno no sabe nada*" (E04, 166-169). Esta voluntad de ser se traduce en un compromiso con la función docente, y lo que se advierte en la práctica es precisamente "*una falta de, quizás, de compromiso con lo que (es) el proceso de enseñanza aprendizaje*" (E01, 112-114).

Otra actitud, ya insinuada es la "vocación" pero en cuanto se encarna en prácticas determinadas. La vocación es un concepto complejo y bajo él se espera que "*sea el profesor con vocación es uno que quiere transmitir los conocimientos y las buenas cosas de la vida y el preparar para que hagan, como le digo, este mundo mejor*" (E03, 225-229), y que tenga entonces "*una fuerte motivación de enseñar*" (E03, 373-374). Las tres características de la vocación que anota el entrevistado: transmitir conocimientos, transmitir las buenas cosas de la vida, y preparar para transformar el mundo, no concitan el consenso. De hecho, un docente expresa que en su opinión "*no está en los profesor universitarios formar una sociedad más humana; yo creo que mayoritariamente para ellos está formar profesionales*" (E04, 277-281). Esto pone sobre la mesa la discusión acerca del grado de amplitud de la profesión docente o, en otras palabras, hasta dónde el profesor es profesor y cuándo comienza a ser una persona común y corriente: ¿cuál es el límite entre lo público (profesión) y lo privado (vida personal)?

6.3. Las actitudes de los docentes hacia los estudiantes.

Estas actitudes se enmarcan en la dimensión relacional de la docencia, que es el eje más constitutivo de la misma: la enseñanza es un acto interpersonal, necesariamente y su esencia es el encuentro de las personas en una situación, una dinámica y una teleología determinadas.

Las siguientes preguntas orientan nuestro análisis: ¿Cuáles son, en opinión de los propios docentes, aquellas actitudes que ellos deberían mantener hacia los estudiantes? ¿Cuáles son las actitudes menos recomendables?

Proximidad es un término clave: próximos hasta el grado de "*involucrarse con los alumnos de alguna manera, ¿cierto?, comprender dónde está el problema en el aprendizaje...*" (E02, 55-57). Se constata también que en la institución ha habido progresos en ese sentido, "*la Universidad de N.N. ha cambiado, noto una mayor calidez, noto una mayor aproximación a los estudiantes*" (E04, 161-163).

La ruptura de la dura verticalidad de la relación docente-discente propia de épocas más autoritarias implica que a los estudiantes se debe "*tratarlos con esa consideración*" (E08, 245) que se le otorga a la autoridad superior. En este sentido, ser considerado es tratar a otro como un auténtico otro, como un yo real, auténtico y valioso, sin condicionarlo ni forzarlo. Para ello "*nosotros debemos estar abiertos todo el tiempo a escuchar a los jóvenes, a la comprensión de lo que estamos tratando de iniciar*" (E09, 225-228).

Otra importante actitud en la relación docente con los estudiantes tiene que ver con el rol orientador del profesor, "*un encauzador, un orientador, una persona que más bien enseña a, ¿cómo se dice?... a aprender y no enseña directamente*" (E06, 193-195). En una palabra, "*nosotros debemos estar disponibles*" (E08, 231).

Por otra parte, el más importante correlato negativo de estas actitudes será la indiferencia, la que expresa el docente como "*se le paga para que haga clases y punto. Eso es lo que hace y simplemente eso es lo que va a hacer*" (E02, 60-63). La distancia o indiferencia ha sido una característica difundida en los medios universitarios por años. "*Hasta hace muy poco tiempo los profesores universitarios eran distantes, (...) personas que, con una imagen de sí mismos de "el que sabe, tiene el conocimiento, el que tiene la información*" (E04, 145-150). La constatación de otro docente es que "*en términos generales, yo creo que el profesor universitario mantiene una distancia con sus alumnos*" (E06, 132-134), lo cual lleva finalmente a esta misma distancia "*deshumaniza el proceso; o sea, si no hay un contacto del profesor directo con sus alumnos, en el cual pueda entenderlos, en el cual pueda interactuar con él, pueda guiarlo, indudablemente se pierde la parte humana del proceso de enseñanza*" (E06, 139-143).

6.4. Las actitudes de los docentes hacia la tarea de enseñar.

Entendemos que no sólo las actitudes hacia los estudiantes son relevantes sino que hacia la propia tarea de enseñar, máxime cuando se trata de personas que ejercen una profesión para la cual no han sido preparados formalmente, sino que la han asumido a partir de sus calificaciones académicas.

En la cultura universitaria chilena la docencia es asumida como un componente de un conjunto más amplio de funciones, la denominada "carga" académica. Ahora bien, ha existido en el país en los últimos 30 años un énfasis crecientemente en la investigación, lo que ha estratificado de alguna manera estas funciones, poniendo a la investigación en un rango superior a la docencia. Si a esta situación desmedrada le sumamos el hecho de que la mayoría de los docentes universitarios no son profesores sino que provienen de otras profesiones, entonces el referente para ellos será su experiencia anterior como estudiantes, sumada a sus expectativas actuales como investigadores.

Repetir esquemas es una estrategia con la cual los docentes universitarios se aproximan al ejercicio profesional. Muchos de ellos intentan reproducir esquemas que en su opinión han resultado exitosos en el pasado: "*lo que ellos consideran buenos profesores para ellos, quieren serlo para sus estudiantes*" (E01, 422-424). Entonces, ante la carencia de preparación profesional, actúan replicando estilos, criterios, formas de exposición, estrategias de evaluación. Claro está que con el correr de los años estas formas de acción van cristalizando, se van estableciendo como patrones casi paradigmáticos, de los cuales será muy difícil escapar.

Sin embargo, también encontramos docentes que tienden a repetir modelos porque no tienen tiempo o carecen de interés: "*docentes universitarios como que repiten estas metodologías históricamente o porque no tienen tiempo o, a veces también, porque no tienen interés en cambiarlas, o sea, están cómodos en eso, significa trabajo cambiarlas, o no quieren enfrentarse a lo mejor, a una situación en la cual puedan verse aminorados en su prestigio*" (E06, 297-305).

Por otra parte encontramos que aparece el tema de la motivación hacia la profesión, "*una fuerte motivación de enseñar y no un profesor que se está ganando un sueldo*" (E03, 373-375), de forma que el profesor universitario puede "*estar feliz de enseñar, el considerar (que) tengo la mejor profesión del mundo*" (E03, 380-382). Por cierto una experiencia exitosa del acto de enseñar es aquella en la cual el profesor "ve" cómo el alumno aprende, situación que es completamente gratificante. En palabras de un docente "*es tan rico cómo uno tiene ese... cómo se da esa chispita especial a los alumnos, o sea, es mágico*" (E05, 338-340).

Sin embargo también existe la otra cara: la desmotivación. Una crítica que se hace es que los "*profesores están adoleciendo de motivación, de capacidad para incentivar, para estimular, para hacer que el alumno o los estudiantes se sientan interesados en aprender*" (E11, 179-183).

Finalmente, aparece la profesión de enseñar como un rol modelar: "*ir transmitiendo esas cosas es a través del ejemplo, no creo mucho que se haga a través del discurso*" (E08, 214-216), entendiendo que lo que se enseña debe estar encarnado en la propia existencia del enseñante. Si enseña la honradez como un valor en los negocios entonces debe ser honrado en sus propios negocios. Si enseña la importancia de ser crítico, entonces debe ser crítico con lo que sabe y aceptar la crítica. Esta consistencia es la que permite finalmente que el profesor sea considerado un modelo, "*una especie de paradigma en sus alumnos, cosa que sea un modelo a imitar, ojalá*" (E06, 220-222).

6.5. A modo de síntesis.

Los resultados encontrados se resumen en las siguientes expresiones.

- Al referirse a las actitudes esperadas de los docentes, hay una tendencia a una identificación romántica del profesor, como añoranza del maestro o la maestra de primaria;
- Diversas actitudes son esperadas en y por los docentes: preocupación por los aspectos humanos, "no académicos"; orgullo y amor por la profesión docente; alegría de enseñar; ética y espiritualidad; vocación por la enseñanza como eje vital; hacerse responsable del estudiante; criterio flexible, elástico y adaptativo;
- Las actitudes observadas contraponen la imagen ideal (esperada) con la prosaica preocupación por ganar dinero, siendo la enseñanza un trabajo más, no una vocación; la falta de compromiso con el aprendizaje es una consecuencia de ello;
- Las actitudes esperadas hacia los estudiantes tienen que ver con la proximidad, la apertura, la disponibilidad. Alternativamente se observa distancia, autoritarismo, deshumanización de la relación profesor-alumno.
- En referencia a la tarea de enseñar, la actitud más habitual es la de repetir los esquemas que, supuestamente, han sido exitosos con ellos mismos durante su proceso de formación en el pre o postgrado. También se constata una fuerte desmotivación por la enseñanza.

7. Conclusiones y discusión.

En el texto se ha realizado una exploración mediante expresiones, opiniones, juicios de valor, relatos de experiencias, provenientes de quince docentes universitarios quienes dispusieron de su tiempo y reflexión para aportar a este trabajo.

Se han revisado cuatro acápites fundamentales: aprendizaje, enseñanza, conocimientos previos, actitudes docentes. En las siguientes líneas se sintetizan los hallazgos en cada una de las dimensiones, y luego se discuten brevemente a la luz de la experiencia de la misma universidad en que fueron hechas las obaservaciones.

7.1. Conclusiones en relación a la concepción de aprendizaje.

En relación a la concepción del aprendizaje podemos señalar que en general se sostiene una visión más próxima al conductismo, con fuerte peso de la pedagogía basada en objetivos conductuales. Hay también algunas aproximaciones al modelo del cognitivismo más bien piagetano, especialmente a través de la metáfora del procesador de datos. También se encuentra que los docentes asocian la modalidad esperada de aprendizaje con las características epistémica de la disciplina que enseñan. De esta forma, la naturaleza (semántica, sintáctica) de la disciplina es la que rige la forma del aprendizaje de la misma.

La relación profesor-alumno es valorada y apreciada por los docentes, aunque siempre en una asimetría que mantiene el rol directivo del docente, concentrando el poder y el control. Los docentes no conciben o, al menos, no lo expresan, el trabajo cooperativo o colaborativo como una estrategia valiosa y válida de aprendizaje para sus estudiantes. El aprendizaje entonces es concebido como un acto individual, solitario. En el plazo docente, esta asimetría se manifiesta en la concepción de "entrega" o "transmisión" de saberes con el propósito de "rescate" de los estudiantes del estado de "conocimiento cero" en que se asume que están.

Se valoran los aprendizajes relacionados con conocimiento simple, aplicación de algoritmos, resolución de problemas. También hay una valoración de las funciones de comprensión, especialmente respecto de un concepto amplio de cultura o mundo. Si bien se concibe el aprendizaje como un cambio, como un paso, el alumno es concebido más como sujeto al que le acontecen los cambios que como el que los produce.

7.2. Conclusiones en relación al concepto de enseñanza.

Respecto a cómo conciben la enseñanza los docentes, hay una fuerte tendencia a significarla como un proceso de transmisión de contenidos de diverso tipo, orientada a la producción de cambios conductuales en los estudiantes. Estos cambios pueden ser de variadas formas: desde repetir unidades de información nuevas, a aplicar algoritmos o cuasi-algoritmos en la resolución de problemas más o menos tipo, hasta la pretensión de que los estudiantes desarrollen conductas complejas de comprensión, análisis crítico y reflexión metacognitiva.

La relación profesor-alumno en la perpectiva de la enseñanza se mantiene siempre en la asimetría, donde el profesor concentra el poder en sus más variadas expresiones: el poder del conocimiento, el poder aprobatorio, el poder sancionador. Al considerar al estudiante meramente en su dimensión de individuo y no como miembro un grupo más complejo que la pura suma de las individualidades, se hace una reducción de la visión de la persona del alumno, y se lo trata aisladamente. Las comparaciones con los demás estudiantes tienen sentido solamente para determinar quién es más o quién es menos que otros.

Hay conciencia de la falta de preparación didáctica de los docentes, pero también es cierto que abundan las estrategias de enseñanza, se citan las de nivelación de grupos; agrupación en función de niveles de dominio; experimentación didáctica; orientación tanto por motivación como por "coaching"; búsqueda de significación; establecimiento de contratos; aprender haciendo. En general son estrategias de carácter intuitivo, desarrolladas sin rigurosidad técnica ni menos aún con un monitoreo sistemático ni un proceso reflexivo que las acompañara construyendo conocimiento sobre las propias prácticas. Este aspecto "reflexivo" de la docencia (en el sentido del profesional reflexivo de D. Schön) está completamente ausente de las expresiones de los profesores universitarios entrevistados.

La enseñanza no es una tarea fácil, y los docentes manifiestan que el excesivo número de alumnos así como las restricciones en el tiempo de la docencia constituyen fuentes de dificultad para el ejercicio de su trabajo.

7.3. Conclusiones en relación a los conocimientos previos de los estudiantes.

La idea de los conocimientos previos es análoga a las denominadas "conductas de entrada" en la tradición de la enseñanza basada en objetivos. El concepto de conocimiento previo, siendo más amplio que las conductas de entrada funcionales a las demandas de la instrucción en un nuevo período o estadio de la misma, es asumido por los docentes precisamente como sinónimo. Por consiguiente, encontramos entre los docentes universitarios una noción de conocimiento previo como conocimiento funcional para el inicio de la enseñanza y el aprendizaje en una materia determinada. No se consideran en este concepto otros conocimientos asociados, como tampoco las visiones ni, menos aún, las variaciones que obedecen a peculiaridades de las diferentes subculturas que constituyen el bagaje cultural de los estudiantes.

Manteniendo la prevención anterior, hay opiniones diversas sobre la importancia de evaluar los conocimientos previos de los estudiantes. Hay quienes dicen hacerlo y utilizar los resultados para determinar un punto de partida, aunque el tiempo que destinan al proceso mismo del examen es completamente ínfimo para el propósito (quince minutos al inicio del semestre). Hay también quienes declaran hacerlo pero no tener idea del para qué de esa evaluación.

Asimismo, no queda claro si la evaluación de conocimientos previos tiene algún impacto sobre la enseñanza (selección de contenidos, secuenciación de la instrucción, ritmo de trabajo, tipo y diversidad de la ejercitación, vinculación con la realidad y la profesión, etc).

Los conocimientos previos constituyen un hito referencial tanto respecto de los estudiantes de primer año como de cursos superiores. Hay una doble distorsión en relación a los alumnos de primer año; por una parte, se desconocen sus características, conocimientos previos, historias; en otras palabras, se les hace invisibles. Por la otra, al partir de la base de que nada saben, se desvaloriza completamente lo que pudieran saber; pero como nunca se sabrá, nunca podrá ser valorado. Para los alumnos de cursos superiores se plantean también problemas de conocimientos previos insuficientes, pero ahora el referente no es la enseñanza media sino que la propia universidad. Finalmente cabe destacar que la evaluación, cuando se hace, se lleva a cabo de manera muy poco técnica, más bien intuitiva.

7.4. Conclusiones en relación a las actitudes de los docentes.

Los docentes evidencian una tendencia a la identificación romántica del profesor, como añoranza del maestro o la maestra de primaria. En esta visión lejana, desdibujada, aparecen algunos rasgos que serían las actitudes esperadas en y por los docentes: la preocupación por y prioridad de los aspectos humanos, "no académicos"; el orgullo y amor por la profesión docente; la alegría de enseñar; la dimensión ética y de espiritualidad; la vocación por la enseñanza como eje de la propia existencia; el hacerse responsable del estudiante; tener un criterio flexible, elástico y adaptativo. Estas disposiciones son apreciadas por los docentes. No obstante, constituyen una imagen ideal, no apropiada para los tiempos y condiciones actuales.

Por lo anterior, las actitudes observadas contraponen la imagen ideal (esperada) con la prosaica preocupación por ganar dinero, consistiendo la enseñanza en un trabajo más, como cualquier otro y no una vocación en primer lugar. Como consecuencia, aparece la falta de compromiso con el aprendizaje de los estudiantes, su tratamiento como "cliente" de segunda categoría, aquel que no tiene donde elegir; en términos de mercadeo, una clientela cautiva.

Las actitudes esperadas hacia los estudiantes tienen que ver con la proximidad, apertura y disponibilidad. Todas actitudes valoradas como importantes. Sin embargo, alternativamente se observa distancia, autoritarismo, deshumanización de la relación profesor-alumno.

En cuanto a la tarea de enseñar, la actitud más habitual es la de repetir los esquemas que, supuestamente, han sido exitosos con ellos mismos durante su proceso de formación en el pre o postgrado. También se constata una fuerte desmotivación por la enseñanza.

7.5. Discusión.

Los hallazgos que se informan ponen de manifiesto una serie de temáticas que no han sido abordadas de manera frontal en el sistema educacional chileno. Específicamente en el ámbito de los tópicos relacionados con la docencia, existen cuestiones de relevancia tanto interna para la universidad, como para el sistema educacional en su conjunto. Entre ellas destaca la relación entre la enseñanza media y la universitaria (como articulación y como transición), la formación de los docentes universitarios, la organización del currículo de formación en el pregrado, la posición o jerarquía de la docencia en la estructura de la tarea académica.

Cambio y autonomía.

El carácter autónomo de las corporaciones universitarias hace imposible que se establezcan regulaciones por el Estado sobre las prácticas docentes. La autonomía permite la libertad más completa en relación a qué, cómo y cuándo enseñar. La regulación no ha de venir del Estado, hacia el cual existe una desconfianza radical, sino que del mercado que se asume como el mejor asignador de recursos. La ideología que sustenta al sistema universitario es la que se implanta por el gobierno militar desde 1981 y consagrada en la Ley Orgánica Constitucional de Enseñanza, dictada por el mismo gobierno y que tras de doce años todavía no puede ser reformada por carecerse del quorum calificado que requiere para su tramitación.

Sin embargo, existen instrumentos del Estado por los que puede ejercer (y de hecho ejerce) una presión indirecta; entre estos, los fondos concursables para investigación, los aportes diferenciales según indicadores de eficiencia académica, los proyectos de mejoramiento de la educación superior. Lo característico de estos instrumentos es que son de carácter financiero: el Estado condiciona la entrega de recursos al cumplimiento de ciertas metas o estándares.

Relación enseñanza media y universitaria.
El tema de la relación entre la enseñanza media y la universitaria ha sido discutido con frecuencia pero no se ha logrado resolver los problemas que le afectan. Entre éstos, la articulación es uno de los más trascendentes. De hecho, no existe relación formal entre el currículo de la enseñanza media y el del primer año universitario. Los programas universitarios hacen caso omiso de los programas de formación de nivel medio. Con ello no sólo se genera una brecha de dimensiones desconocidas entre ambos currículos sino que desde la universidad representa el desconocimiento y desvalorización de lo que ocurre en la educación media.

Visto desde la óptica de los estudiantes, su transición de la enseñanza media a la universitaria tampoco es considerada de manera particular, si bien en los últimos años ha existido una mayor preocupación al menos en la institución en la que se hace la investigación. El desconocimiento de los logros de los estudiantes les desvaloriza en cuanto personas y también los homogeneiza, haciendo invisibles sus características peculiares y transformándolos en un promedio, una abstracción completamente irreal.

Formación pedagógica de los docentes universitarios.
La formación de los docentes universitarios desde el punto de vista pedagógico es una cuestión crucial, máxime si prácticamente la totalidad de ellos no tienen formación sistemática en el campo pedagógico. Existen diversas iniciativas a través de programas de formación pedagógica para docentes universitarios (diplomas o maestrías); sin embargo, su modalidad, en general completamente similar al formato tradicional de la docencia universitaria, hace difícil la transferencia al aula, por cuanto no ha logrado impactar las prácticas pedagógicas.

Tampoco ha existido un proceso de acompañamiento o asesoramiento de los docentes que han cursado dichos programas de perfeccionamiento, de manera que si hubo algún "aprendizaje" su falta de monitoreo hace imposible rastrear el impacto de dichos programas. La universidad del estudio ha puesto el cursar el programa de pedagogía universitaria como un requisito para ascender en la jerarquía académica, pero sólo para los niveles más bajos; de esta forma, quienes ocupan las más altas jerarquías están exentos de dicha obligación.

Organización del currículo de formación en el pregrado.
El currículo de formación profesional en el contexto de la universidad en estudio se caracteriza por su rigidez, no existiendo para los estudiantes la posibilidad de elegir entre diversas orientaciones al interior de sus carreras. De esta manera, la rigidez se convierte en una propuesta de homogeneidad para los egresados y para la sociedad. A la vez, pone sobre la institución y los docentes la presión de enseñar "toda" la profesión en un marco de tiempo completamente insuficiente. De ser así las cosas, entonces al menos puede esperarse que cada una de las piezas de este currículo engrane completa y perfectamente con las restantes, en particular con las vecinas. Es decir, se espera una adecuada articulación del currículo. No obstante, es preciso considerar si la propuesta implicada de la plena articulación tiene sentido en el marco de la formación profesional, no sólo por consideraciones relativas a la libertad de elegir una orientación por parte de los estudiantes, sino que fundamentalmente por la dinámica de las profesiones en el mundo real. Todas las profesiones están sufriendo procesos transformacionales que las hacen tecnológicamente más dependientes, las especializan hacia formatos más restringidos, se abren a campos antes no concebidos. De esta forma, queda pendiente la pregunta por la deseabilidad de esta articulación, y alternativamente se abre la pregunta por las ventajas que ofrecería un currículo más abierto, más dinámico, con mayores opciones.

De las opiniones vertidas por los docentes se deja entrever una cierta falta de coordinación al interior del currículo de pregrado que tiene que ver también con temas de articulación y transición. La articulación se relaciona con la forma en que una asignatura engrana con otra

Posición o jerarquía de la docencia en la estructura de la tarea académica.
La enseñanza en una función fundamental y propia de la tarea académica: este es un principio sobre el cual no cabría discusión. Sin embargo, otro asunto es la jerarquía o posición que la docencia ocupa al momento de estructurar la tarea académica de cada docente o de una unidad completa. En este momento irrumpen otras demandas, todas académicas, como investigación, publicaciones, asistencia técnica, gestión, extensión cultural. Una aproximación racional es la que utiliza la universidad del estudio, y consiste en establecer rangos de dedicación a la docencia, y restantes funciones. Sin embargo, al mismo tiempo los criterios para calificar el desempeño, así como para determinar la promoción en la jerarquía académica, favorecen los antecedentes en materia de investigación y publicaciones. La docencia prácticamente no tiene peso, con lo cual queda relegada a un segundo plano. Esto hace que los académicos tiendan a tener menos docencia, a desplazar la tarea docente hacia el personal de contrato parcial o por hora, y a desatender esta función que ha sido declarada tan fundamental.

Recuperar la posición de la docencia en la jerarquía de las preferencias académicas es una tarea importante, y requiere tanto un discurso apropiado, como señales claras acerca de cuánto y cómo será considerada. A la vez, se requiere un proceso formativo que no se plantee sólo en los términos de la enseñanza tradicional, sino que constituya un sistema de auténtico acompañamiento y asesoramiento eficaz para los docentes.

Lo anterior demanda un cambio cultural en el ámbito universitario. Se requiere una "nueva cultura" docente (Imbernón, 1998), una cultura en que se superen las prácticas que hacen del aprendizaje y la enseñanza procesos difíciles, ingratos, y poco motivadores. El argumento más radical es el relativo al individualismo que reina en el medio académico. Se necesita con mucha fuerza superar el aislamiento de y en las aulas: aislamiento de un aula con otra, de una asignatura respecto de la otra, de una disciplina frente a la otra; y aislamiento en el aula, tal que cuando se cierra la puerta comienza un proceso completamente aislado del conjunto, completamente incontrolable y cerrado. Ciertamente el aislamiento no es sólo físico, también es curricular, disciplinario, intelectual, profesional/estudiantil, vital. La formación profesional, en este marco, no puede ser solamente orientada al individuo, sino que precisa promover competencias complejas relacionadas con la vida en común, con la vida ciudadana, con el trabajo colaborativo, con el establecimiento de redes tanto profesionales como amicales.

Por cierto, no es posible caer en la misma situación que asume cero conocimiento en los estudiantes, invalidándolos y haciéndolos invisibles. Existe en los docentes universitarios una visión de la docencia y una experiencia acumulada muy valiosa. Es preciso conocerla más en profundidad, más aún de lo logrado en este estudio. De esta forma se podrán diseñar apropiadas estrategias para fortalecer una función tan relevante y trascendente para la vida de tantas personas. El rescate de estos saberes –y la consiguiente legitimación de los mismos- para su análisis crítico por parte de los mismos docentes puede ser un camino interesante por donde iniciar procesos que lleven a cambios favorecedores del aprendizaje de estudiantes y maestros.

Notas

1. En sentido lakatosiano de la expresión "programa de investigación"

2. Esto es muy notable en la revisión de la base de datos de ERIC Clearinghouse.

3. Anualmente la Vicerrectoría Académica de la Universidad abre un concurso para Proyectos de Innovación Docente, a los que postulan académicos de las diferentes unidades.

4. Hay otras acepciones del término como, por ejemplo, la que plantean Arancibia, Herrera y Strasser en referencia a los procesos de perfeccionamiento de profesores: "*consiste en pequeños grupos de profesores que trabajan juntos, se observan mutuamente, se dan sugerencias, y comparten su experiencia a diario*" (2000:238).

5. Expresión popular para designar "muchacho".

Referencias

Arancibia, Violeta; Herrera, Paulina; Strasser, Katherine. 2000. *Manual de Psicología Educacional*. Santiago: Ediciones Universidad Católica de Chile.

Baena Cuadrado, Ma. Dolores. 1995. *Tareas académicas y teorías implícitas del profesorado. Estudio de casos en enseñanza de las* ciencias. Tesis Doctoral. Universidad de La Laguna, Centro Superior de la Educación, Departamento de Didáctica e Investigación Educativa y del Comportamiento.

Ball, Stephen. 1987. *La micropolítica en la escuela. Hacia una teoría de la organización escolar*. Barcelona. Paidós.

Barragán Medero, Fernando. 1988. *Las teorías sexuales infantiles, la información sexual y las teorías implícitas en los adultos sobre sexualidad y educación sexual: bases para el diseño curricular de la educación sexual en el Ciclo Medio de la E.G.*B. Tesis Doctoral. Universidad de La Laguna, Centro Superior de la Educación, Departamento de Didáctica e Investigación Educativa y del Comportamiento.

Bernstein, Basil. 2000. *Hacia una sociología del discurso pedagógico*. Bogotá: Magisterio.

Díaz B., Frida; y Hernández R., Gerardo. 2001. *Estrategias docentes para un aprendizaje significativo*. Bogotá: McGraw-Hill Interamericana.

Díaz Vila, Mario. 2001. *Del discurso pedagógico: problemas críticos. Poder, control y discurso pedagógico*. Bogotá: Magisterio.

Dienes, Zoltan & Perner, Josef. (1999) A Theory of Implicit and Explicit Knowledge. *Behavioral and Brain Sciences* 22 (5): En http://cogsci.so-ton.ac.uk/bbs/Archive/bbs.dienes.html

Echeverría, Rafael 1995. *Ontología del Lenguaje*. Santiago: Dolmen.

Entwistle, Noel. 1997. Reconstituting approaches to learning. A response to Webb. *Higher Education* 33: 213–218, 1997.

Fernández Pérez, Miguel. 1999. *La profesionalización del docente. Perfeccionamiento. Investigación en el aula. Análisis de la práctica*. Madrid: Siglo XII Editores (2ª edición corregida y aumentada).

Greenwald, Anthony G; Banaji, Mahzarin R.; Rudman, Laurie A; Farnham, Shelly D.; Nosek, Brian A. Nosek; Mellott, Deborah S. Mellott. 2001. A unified theory of implicit attitudes, stereotypes, self-esteem,& self-concept. *Draft of January* 6. En http://faculty.washington.edu/agg/pdf/Unified.010601.pdf

Hativa, Nira. 1998. Lack of clarity in university teaching: A case study. En *Higher Education* 36: 353–381, 1998.

Imbernón, Francisco. 1998. *La formación y el desarrollo profesional del profesorado. Hacia una nueva cultura profesional*. Barcelona: Graó, 1998 (5ª edición).

Jiménez Llanos, Ana Beatriz y Cabrera Pérez, Lidia. 1999. Aproximación a las teorías implícitas del profesorado de educación infantil y primaria, secundaria y superior sobre los medios de enseñanza. En *Revista Pixel-Bit*, 1999, Nro 13.

Lucarelli, Elisa. 2002. Teoría-Práctica, ¿problema eje de la innovación en la enseñanza universitaria? En *La 6 ème BIENNALE de l'Education et de la Formation*, Paris: Mercredi 6 Juillet.

MacFarlane, 1999. Universities in a Knowledge Economy. En Smith y Langslow, 1999, *The Idea of a University*, Jessica Kingsley, pp.124-147

Marrero Acosta, Javier. 1988. *Teorías implícitas y planificación del profesor*. Tesis Doctoral. Universidad de La Laguna, Centro Superior de la Educación, Departamento de Didáctica e Investigación Educativa y del Comportamiento.

Martínez Licona, José Francisco. 1995. *Las teorías implícitas y la resolución de problemas en la enseñanza: el pensamiento de profesores en formación inicial: Las teorías implícitas y la toma de decisiones sobre situaciones problemáticas*. Tesis Doctoral. Universidad de La Laguna, Centro Superior de la Educación, Departamento de Didáctica e Investigación Educativa y del Comportamiento.

Riquelme Bravo, Paula. 2000. *Teorías Implícitas y su influencia en la práctica pedagógica de un grupo de profesores pertenecientes a una escuela urbano marginal de Temuco*. Santiago, Congreso Nacional Reduc "Investigación Educativa e Información", 16-17 Octubre 2000.

Trigwell, Keith; Prosser, Michael y Waterhouse, Fiona. 1999. Relations between teachers' approaches to teaching and students' approaches to learning. En *Higher Education* **37**: 57–70, 1999.

Smith, David. 1999. The Changing Idea of a University. En Smith y Langslow, 1999, *The Idea of a University*, Jessica Kingsley, pp.148-174

Van Nieuwkerk, Karin. 1999. "Un'ora per Dio e un'ora per il cuore": Islam, genere e spettacolo femminile in Egitto. En *Music and Anthropology*, Nro 3.

Vande Walle, Don (2001). The Attributional Leadership Model Reconsidered: The Influence of Implicit Theories on Leader Attributions and Behavior. This paper was presented as a part of the Symposium *Social-Cognitive Approaches to Personality and Motivation in Organizational Behavior* at the 2001 Academy of Management National Meeting, Washington, D.C.

Zurita, Reginaldo. 1992. *La docencia universitaria y los Ciclos Básicos en Chile*. Santiago: Corporación de Promoción Universitaria (CPU).

Acerba de los Autores

Gustavo Hawes Barrios
Sebastián Donoso Díaz
Académicos del Instituto de Investigación y Desarrollo Educacional de la Universidad de Talca–Chile
Teléfono 56-71-200253
Email: ghawes@utalca.cl; sdonoso@utalca.cl.

Gustavo Hawes Barrios, Académico del Instituto de Investigación y Desarrollo Educacional de la Universidad de Talca (Chile) y Profesor del área de Epistemología Aplicada del Programa de Magíster en Política y Gestión Educacional de la Universidad.

Sebastián Donoso Díaz, Académico del Instituto de Investigación y Desarrollo Educacional de la Universidad de Talca (Chile) y Profesor del área de Política Educacional o del Programa de Magíster en Política y Gestión Educacional de la Universidad.

Este estudiuo es parte del proyecto Mecesup TAL0101.

Copyright 2002 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

This article has been retrieved 3722 times since March 27, 2003

Education Policy Analysis Archives
Volume 11 Number 12 March 27, 2003 ISSN 1068-2341

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article if EPAA is credited and copies are not sold. EPAA is a project of the Education Policy Studies Laboratory.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Teacher learning in context: The special case of rural high school teachers

Jay Paredes Scribner
University of Missouri-Columbia

Citation: Scribner, J. P. (March 28, 2003). Teacher learning in context: The special case of rural high school teachers, *Education Policy Analysis Archives, 11*(12). Retrieved [date] from http://epaa.asu.edu/epaa/v11n12/.

Abstract

Falling under the umbrella of teacher quality, professional development is an important policy issue in US public education. Understanding teacher learning and its relationship to teacher work is critical if efforts to improve teacher quality are to be successful. This article examines one overlooked context in the discourse about teacher learning and work—rural high schools. The study focuses on 20 teachers across 3 case study schools and conceptualizes the relationship between teacher learning and work according to three contexts: the core, intermediate and peripheral contexts. These contexts are explored and important features discussed.

A growing consensus exists among scholars, policymakers, and education leaders that sound organization and implementation of professional development enhances teacher quality and, as a result, student learning. For instance, it has been argued that professional development resources should focus on (1) closing the gap between teachers' knowledge and student performance goals (Hawley & Valli, 1999; Sykes, 1999), (2) strengthening teachers' understanding of the connections between content and students' thinking and learning (Thompson & Zeuli, 1999), and (3) creating a tighter organizational fit between teacher learning activities and teacher work (Little, 1999). In spite of this growing consensus, traditional approaches to professional development persist in most schools. This study continues a line of research (Scribner, 1999) that aims to more deeply understand the reasons behind this persistence. Specifically, the study explores the relationship between teacher work context and teacher learning in rural high schools. In so doing, the study describes from the perspectives of teachers a broad view of work context that spans across the institution of education, with each contextual level influencing teacher learning in different ways.

In short, rural schools provide an important context for study for several reasons. Foremost, rural schools have been historically under studied. But also, exploring teacher learning in rural contexts provides a platform from which to question assumptions inherent in prevailing conceptions of effective professional development policy and practice, most of which emanate from research that focuses on urban and suburban teachers and students. The following research questions were addressed:

- What are the predominant features of work context according to these teachers?
- How does the work context of rural high school teachers influence how they learn?

In the discussion section I explore the relationships among the themes developed and draw implications for policy, practice, and future research.

Conceptualizing Teachers' Work Context

Describing the nature of teacher work has been the focus of more than a few researchers (Eraut, 1994; Hargreaves & Goodson, 1996; Huberman, 1993; Jackson 1986; Johnson, 1990; McLaughlin, 1993; Scribner, 1999; Talbert & McLaughlin, 1996). While the purpose of this study was to further develop a grounded theory to explain the relationship between work context and teacher learning, it is useful to examine how others have conceptualized teacher work. Johnson (1990) conceptualized teacher work as being comprised of multiple dimensions including, political, economic, physical, organizational, psychological, cultural, and sociological. As such, Johnson's "constellation of workplace variables" described the workplace as a place where structure of formal authority, organizational policies and procedures, and informal norms that shape behaviors, beliefs, and actions converge. To further complicate matters, the manner with which these variables converge in any given school or district varies across organizations. What is fairly constant, Johnson argues, is the *presence*—in some form—of these variables.

Two general perspectives of the relationship between teachers and their work context obtained in the literature. One view took a deterministic slant—that teachers are almost fully constrained by their work context. For example, Hatton (Hatton, 1987) argued that "situational constraints" inherent in teachers' workplaces exist in the present and past to shape teachers' dispositions, behaviors, and actions taken in the here-and-now. One such constraint, resources, can support a status quo approach to practice in cases where the resources to support innovative ideas do not exist. A more subtle constraint, hidden pedagogies (Denscombe, 1980, 1982; Hatton, 1987), reflects the cultural and historical dimensions of school contexts that contribute "significantly to the formation of culturally based attitudes, preference, and dispositions which have their own momentum" (Hatton, 1987).

Others, however, argued that while organizational constraints might influence the nature of teacher work, teachers maintain a certain freedom to act autonomously within the constraints, or presumably to expand the limits of those constraints (Grant & Sleeter, 1987). Grant and Sleeter argued that "teachers *do* have room to act, and they do not all act in the same way, in spite of similarities in their present and prior experiences" (p. 62). In other words, the fact that teachers working in similar work environments with identical constraints can act in different ways challenges the hegemonic view of teacher work context and its relationship to teacher action. Ultimately, rather than posing an opposite view of the deterministic perspective of teacher work, Grant and Sleeter pose a middle ground argument. That is, teacher action is "neither totally free nor totally determined" (p. 62). Instead, their autonomy lies in a gray area in which "actors draw on practical knowledge to guide their actions, often without acknowledging many of the taken-for-granted conditions that give rise to this practical knowledge, or the unintended consequences of their actions which tend to reproduce those conditions" (p. 62). This middle ground approach supports findings by Scribner (Scribner, 1999) that teachers experience their professional learning broadly, but that work context can shape the possibilities of teacher learning in subtle and not-so-subtle ways (Eraut, 1994; Jackson, 1968, 1986).

Two of the most immediate contextual factors facing teachers within schools are students and teachers' subject matter. Most would agree with McLaughlin (1993) that students are the most prominent feature of the school as workplace. She argued that students help to define the context of teacher work through their diversity, their individual and cohort personalities, their and their families' commitment to school, issues related to school safety, and so on. McLaughlin also found that teachers responded to students in one of three ways: 1) maintaining traditional standards; 2) lowering expectations for coverage and achievement; or 3) adapting practices and pedagogy. While teachers' perceptions of their students influenced how teachers approached their work, a school's mission, organizational structure, formal polices and "patterns of communication" (p. 89) also shaped teachers' objectives toward teaching. However, in high schools the contextual factor of most significance in determining the nature of professional community was the department (McLaughlin, 1993; Talbert & McLaughlin, 1996) where teacher work processes shaped how they perceived their students' abilities, their subject matter, and their relations with other teachers.

Eraut (Eraut, 1994; Scribner, 1999; Talbert & McLaughlin, 1996) provided a framework useful for understanding how the contexts of teacher work might contribute to teacher learning. The classroom context—where teachers spend the bulk of their professional lives—"corresponds with more private contexts in which normal professional practice is produced in a relatively routine manner without questioning assumptions on which it is based" (p. 26). Eraut saw a strong link between classroom contexts and teacher learning. He argued that, "teachers are in a 'doing' environment more than a 'knowing' environment" (p. 31), and therefore, tend to rely on procedural (how to) knowledge that often is acquired unreflectively. In other words, knowledge acquired is not atheoretical, but theories that guide practice remain implicit and serve to make life tolerable. In this regard, Denscombe (Denscombe, 1980) warned that when guided by implicit theories teachers are more susceptible to the displacement of one set of teaching goals for another—i.e., shifting focus to controlling the classroom at the expense of student learning. Thus, while important, knowledge that guides practice often remains tacit; and in the isolation of the "classroom, the only significant validator of knowledge are the teachers themselves....the validation is individual rather than collective" (Eraut, 1994, p. 32).

Clearly, the school also serves as an important context of teacher work. Professional learning in the school context focuses primarily on the business of the organization. For example, developing consensus and solidifying teachers' understanding of the school's mission and vision are central learning activities (McLaughlin, 1993). Learning the language of policy (Eraut, 1994) is an ever increasing focus of school learning also. The language of education reform, such as performance assessment, high stakes testing, and other accountability measures, often becomes the focus of teacher learning for at least a couple of reasons. First, teachers must make sense of reforms that are promulgated from "on high." Second, teachers often sense the political dimensions of their work and realize the need to be able to communicate the purposes and potential impacts of the reforms to parents and other education stakeholders (Bredeson & Scribner, 2000). Thus, learning is sought to cope with external demands, not necessarily to expand the pedagogical repertoires or content expertise of teachers. As a result, it is at the school (and district) level that much professional development is associated with formal activities that are well suited for "getting people on board with the policy language" (p. 31).

Finally, teachers work in a more diffuse context, also. Eraut (1994) calls this the academic context, while others have described professional networks such as professional associations as places where teachers acquire knowledge (McLaughlin, 1993; Scribner, 1999). Contrasted with the procedural and tacit knowledge acquired in the classroom, the academic context is where teachers acquire propositional knowledge and where theories are made explicit. A specialized language characterizes knowledge acquired in the academic context, with a high value placed on understanding theories rooted in the disciplines (Eraut, 1994). Furthermore, in this context knowledge is typically assessed formally to measure its acquisition, although determining how it is ultimately used is less sure. While the knowledge acquired in the academic context is arguably detached from practice, Eraut argues that the knowledge and habits of mind acquired in the academic context are not irrelevant to teachers, "…the academic context has norms that support and expect learning to be a lifelong process, that new knowledge will be acquired by all members of the institution, and that new knowledge will be put to good use" (p. 30). Thus, the academic context is where theories are made explicit for critical analysis.

Teacher Learning: Best Practices

In spite of the attention paid to teacher professional development, organizational learning in schools, professional community, and so on, the act and impact of teacher learning remains difficult to observe and even more challenging to measure over the long term. Fullan's (Fullan, 1995) description of the potentials and limits of professional development continues to ring true. Specifically, one dilemma stands out. That is, as we continue to seek ways to foster school improvement we struggle with balancing the learning needs of organizations and the individual professionals within them (Scribner, Hager, & Madrone, in press). A second dilemma underscores the fact that meaningful change ultimately occurs locally. But still, the change agenda is largely determined external to the school, perhaps at the district level, but more likely at the state level. Finally, the focus of what professional development efforts should look like and focus on continues to be debated. For instance, scholars have argued convincingly that the purposes of professional development in any given school or district should be determined by the gap between student achievement goals and actual student performance (Thompson & Zeuli, 1999). However, actually identifying the teacher learning activities that might fill that gap is easier said than done. Defining professional development as a gap implies that teachers have not been achieving adequately themselves. While certainly the case for some teachers, not all teachers are failing. Secondly, research also suggests that teachers have multiple reasons for engaging in professional learning activities that they believe will assist them in becoming more effective with students, but to the external observer may appear to be unhinged from the core issue at hand—teaching and learning (Scribner, 1999).

With persistent challenges like these, the staying power of teacher professional learning as a policy concern suggests that the concept is important to the lives of students and teachers. Furthermore, in spite of the dilemmas discussed above, a consensus on professional learning practices has emerged that can serve as a useful guide. Consensus surrounding teacher learning today focuses on the importance of learning in context and acquiring knowledge that is relevant to one's professional context. Furthermore, it has been argued that the specific focus should be directly linked to student learning (Sykes, 1999). In addition, to make professional development meaningful in a reform environment, deeper coordination between schools and districts must occur to ensure its relevance. Sykes argued that the coordination between schools and districts should ensure that professional development is driven by the "teacher-student learning connection for the selection and design of teacher professional development" (p. 159). Further, teacher learning should focus on deepening teachers' knowledge of the specific content that students are expected to know. And finally, multiple sources of evidence (especially, the assessment of student work) should be integrated into teacher learning experiences.

Thompson and Zeuli (1999) argued that in the face of standards based reforms, professional development for teachers has failed. Interestingly, they argued that because of the nature of teachers and teaching, tinkering with the social and structural arrangements of teacher learning is insufficient. Indeed, policy makers must ensure that the professional development content and pedagogy are appropriate. In their view, "appropriate" professional development assists teachers with developing their own ideas and connections among the materials that students are to learn, understanding the various ways students experience a given content area, and learning how to foster student engagement with the material.

How teachers experience their own learning has also become an important dimension of teacher professional development. Challenging the notion that teachers are merely tinkerers who favor improving around the margins of their expertise, some believe that transformative learning should be the goal of professional development. For instance, Thompson and Zeuli argued that to be meaningful teacher learning activities must provide cognitive dissonance-creating and dissonance-resolving opportunities related to teachers' classroom experiences. This reflexive approach to professional learning should be designed to develop "new conceptual knowledge (understanding), rather than, say, new habits of practice" (Thompson and Zeuli, p. 356). It is through new conceptual knowledge that new practices develop.

Methods

The present study employed qualitative methods and procedures of grounded theory in three rural public high schools in a midwestern state to further explain the relationship between teacher professional development and the context of teacher work at the high school level (Scribner, 1999). The study utilized a multiple case study design (Yin, 1994). Data collection and analysis occurred over two academic years and involved several stages. First, three rural high schools were identified and an initial investigation was made to ensure the schools were indeed rural in nature (e.g., small, dispersed school populations, seasonally-based community economies). Second, using purposive sampling techniques, in-depth interviews were conducted with 20 teachers across content areas (academic and career and technical education) and years of teaching experience. Teacher interviews focused on teachers' perceptions of their work environments, their learning strategies, and the connections between the two. School administrators were interviewed to develop an understanding of each school's organizational goals and philosophy toward teacher professional development. Interviews lasted from 45 minutes to 2 hours and were audio-taped and transcribed. In each school focus group interviews with previously interviewed teachers were conducted to further explore emergent themes that developed from the analysis of initial interviews and observations. Third, observations of teachers at work and in learning activities were also conducted. In each school three teachers were shadowed for one day each; these teachers were later interviewed to gain further insights. In addition, schools were each visited on other occasions to observe such activities as professional development days, faculty meetings, and extracurricular activities to gain a deeper understanding of teacher work. Finally, teachers were provided an opportunity to read a draft of the manuscript. Several teachers provided comments and insights to further hone categories.

Findings

As a result of a constant comparative analysis of the data, this section explores three contexts of teacher work: the core, intermediate, and peripheral contexts. Each context was described according to its dimensions, including how those dimensions facilitate or impede teacher learning. Most importantly, findings showed that teachers defined their work context more in terms of relationships (and the nature of those relationships) and less in terms of physical plant, resources, pace of work on so on. By examining these relationships in terms of dimensions occurring at different contextual levels the reader will see more clearly how context, broadly defined, influences teacher learning.

The Core Context

The "doing" environment of the classroom is represented here as the core context. Like an atom, the core context is difficult to change precisely due to the strength of relationships that define it. Teachers and their relationships between students and teachers' subject areas comprise this triadic relationship. While the intensity of interaction among this triad may provide a rich source of continuously renewing knowledge, it also requires expeditious problem resolution. This need for decisive action makes the core context difficult to penetrate with knowledge from other contexts.

Students as context. Students did more to shape the context of teacher work than any other dimension. Teachers were keenly aware that students shaped their work contexts hourly, daily, by semester, and each year. The fluidity of the student dimension manifested itself in the ever-changing student body. Teachers were aware that on any given day, semester, or year, their students might come to school with different set of needs. A math teacher described how the changing needs of students influenced his learning and classroom practice:

> You learn from year to year that you can't teach different students the same way. And you won't know that until you get involved with the problems, and what you have to be really flexible. In other words, you have a standard way of doing business, but you're going to see the look on their face, the big question mark come across, and if you see a lot of that, then you're probably going to have to back up and start over again and try a different approach.

However, the fluidity of students also created daily learning opportunities for teachers. According to teachers, students constantly, and in ways difficult to predict, questioned teachers regarding subject matter and other issues that expanded teacher understanding of the content. A novice English teacher described the phenomenon:

> The reason I don't like to use tests, like say a Shakespeare unit we just did, every class is very different, and every class is going to have a very different discussion, and they're going to focus on different elements. They're going to be drawn to a certain character or they're more into analyzing the motivations instead of looking at the overall plot, so I don't like to treat every class the same.

This type of student question brought to light the different levels of subject matter knowledge needed to engage students and facilitate learning. As a first year teacher described having in-depth knowledge of the details of her subject matter were key to "running a smooth class," and it was her students that acted as a catalyst for her learning by causing her to reassess and expand her own content knowledge of world history:

> I was teaching students about Hannibal and his elephants back in ancient history, and one of my students goes, 'well did they use camels in warfare?' I really don't know. Let me see if I can't find that out, and I actually was able to find the information that, yes, the Assyrians used camels in warfare, and that told me that, yes, it's good to have the broad picture, but I need to know the details to make it more interesting and be able to answer some of the questions for the kids.

Another predominant teacher learning theme within the student dimension was learning how to control the classroom. Teachers who addressed this phenomenon described a negotiation process that, through trial and error, allowed teachers to control the classroom in order to achieve learning objectives by establishing a good rapport with students. A veteran English teacher was observed relying on this rapport to motivate an otherwise unmotivated section of students one particular day. She told them:

> You tell me, what do you want to learn? Because, you know, you've got to be in here; I'm not cutting you loose any time soon, so what will make sense to you? This is an English class; you know you've got to do some kind of reading, some kind of writing; we've got to communicate with each other. We know this has to be done. What do you want to do? How do you want to apply it?

That afternoon she explained her approach toward her students:

> That really helped because it gave them a little bit more ownership. Because I do find that a lot of the kids, especially the ones who are not college-bound, do not take ownership of their education. And they'll even say, 'I'm sorry I didn't get my homework done.' And I say, 'you don't need to apologize to me. I've already graduated from high school. This does not affect me. This is yours. You should own it; you're the one that you're messing up.'

Another English teacher described how she moved from direct instruction to a more collaborative learning environment in response to student reaction and behavior:

> I've learned to deal with students most from experiences, I guess. When I first started, I was very traditional, direct instruction, whole group, follow the Madeline Hunter method. Then as I learned more about what it is to be a teacher and what I learned from my students and what they needed, then I began to refine methods and to relax a little bit....One thing I resent about teaching is that you have to spend so much time dampening little fires, especially if it's a whole-class instruction model or direct instruction with the whole class. So, hopefully, working with smaller literature groups, more group activities, you can engage the students better.

Subject matter as context. Stein and D'Amico (Stein & D'Amico, 2000)argued that the degree to which a subject area is perceived as a well defined discipline (e.g., mathematics) or a more loosely defined set of knowledge and concepts (e.g., social studies or language arts) influences how teachers understand and approach their subject matter. These data suggested that these differences might also influence how teachers approach their learning. Teachers in the core academic areas tended to have broader notions of professional learning than their career and technical education colleagues. For example, teachers in the arts and career and technical education areas described learning that focused on developing their own technical skills. Little if any discussion revolved around learning issues of pedagogical theory and skill, education reform, or content area knowledge. The focus of these teachers was squarely on developing concrete, "real world" skills to pass on to their students.

Interestingly, academic teachers also voiced concern that relevant learning be concrete and applicable immediately in the classroom. However, when probed further academic teachers described a process in which they banked knowledge gained. That is, while they expressed a desire to focus their learning on practical and immediately relevant knowledge, they actually sought a wide array of knowledge that often was used over a much longer time frame. While it may not be surprising that teachers associate knowledge use with need, teachers were surprised to admit that they used a wide array of knowledge (including e.g., propositional and procedural) to inform their practice. An English teacher explained how in reality learning and the use of knowledge acquired are not linear processes:

> I've been to a lot of different conferences and have met different speakers and gurus in the field. I have a lot of books I've collected from those things and other places that at the time I didn't have a chance to read. But a lot of times, I'm finding I still pull them down, and they are still current and useful as far as how to teach writing.

Like their career and technical colleagues, several academic teachers described the importance of learning the skills associated with their subject area. They believed that their effectiveness with students relied at least partly on their ability to "practice what they preached." One English teacher described the importance of developing her own skills related to the craft she was teaching to students:

> Participant: I do a lot of free writing. A lot of times there'll be some kind of nugget that comes out of it. The writing group that I started this year helped a lot because a lot of times, we'd just sit around and write. Then we'd read it and talk about it a little bit. And I used to be in a couple of all-teacher writing groups. We just started talking about how important it is for us to start a writing, reading, or art group where we all get together.
>
> Interviewer: Among the teachers?
>
> Participant: Yes.
>
> Interviewer: Why was that important for you professionally?

> Participant: I think I was hassling [my colleagues] about the lit magazine, actually. I kept saying, kids really want to see that we can do what we say we can or what we teach. It's really awful to be a writing teacher and never share your writing or an art teacher and never share your art.

On the other hand, some teachers focused their learning more narrowly on their discipline. For example, a math teacher described how his professional learning was focused squarely on developing a deeper understanding of content knowledge. He believed that the topic of math kept him from having to deal with or, for that matter know about, some of the social issues facing students. For example, when probed about how he had learned to work with a growing number of students who were not white, he responded:

> [Math teachers] don't have that problem [teaching and discussing issues of diversity], whereas I've talked to other teachers from other areas, and that's really a challenge for them. And there again, teaching math, you know, that's the reason I stick with math; I don't want to get involved with other issues. I found out if I can avoid that, I have fewer conflicts with students disagreeing or whatever. If we can all talk about algebra, there's not much to argue about. It's pretty cut and dry.

These examples shed light on the ways students and subject matter together define the core context of teacher work. The triadic relationship is complex. The teacher-student relationship can shape teacher learning of subject matter, and the teacher-subject matter relationship can shape teachers approach to students. The fluidity and unpredictability of teacher work in the core context can make *a priori* attempts at teacher learning seem quixotic. Teacher learning in the core context was experiential, usually non-reflective, and lonely with few signs of encroachment from the intermediate and peripheral contexts. It is no wonder that teachers described learning from experience as the most important source of learning. As a result, the learning spurred by students occurred through experience, i.e. trial and error. Furthermore, the focus of the trial and error learning ranged from learning to control students, to deepening and broadening one's content knowledge, to learning about student experiences outside of school and how those experiences influence the classroom.

The Intermediate Context

The intermediate context focused on the dimensions of teacher work that existed within the school, but beyond the core context. The small faculties in these isolated schools shed light on a high school context different than those already reported in the literature. In particular, three dimensions of school context emerged as most salient in the case study schools: strain of multiple roles, faculty relations, and principal leadership.

Strain of multiple roles. While teachers predictably discussed the problems posed by lack of time and other resources, especially as they pertained to teacher learning, these issues were exacerbated by the small size of these schools. In these rural schools teachers took on numerous curricular and extracurricular activities and often taught a wide array of subject area levels. This stretching of teacher responsibility shaped how teachers thought about their own learning. Across each school, teachers explained that they did not spend adequate time in meaningful discussion with their peers because of the "many hats" all teachers wore to meet the academic and social needs of students.

It was not uncommon for teachers in these small rural schools to be in departments that consisted of one to three teachers. One-person departments were quite common, while departments of more than three were rare. As a result many of the teachers interviewed and observed taught multiple "preps"—some teaching four or five different levels within a subject (and across subjects) each day. The daily grind caused by these multiple class preparations had a profound effect on how teachers approached their own learning. Faced with four or five class preparations teachers had to simultaneously cover a wide breadth of subject matter knowledge and varying levels of student needs. For example, during one set of observations an English teacher described how preparing and continuously honing multiple sessions had constrained her ability to deeply explore content knowledge. She described how she only marginally improved her classes and was reluctant to commit too much time to in-depth learning in one area for fear that three or four other groups of students would suffer. Furthermore, because teachers were constantly preparing for a variety of classes they did not have time to discuss issues of practice with other teachers or collaborate on interdisciplinary units. The following exchange during a focus group reflects teachers' frustration:

> Participant 1: We do need to do more. See, we tried last year. Remember, we tried to do the big unit. We were going to do it around the Olympics or something, and it's just we got so busy that it just didn't get done.
>
> Participant 2: I think at the high school level, we need to collaborate more, team-teach more. That's something that this school, we haven't done. But we're all interested in doing that. The problem is that because we're limited with faculty, our planning times are different. After school, we have so many committee meetings, because in a small school district, one person could be in charge of ten committees because we just don't have as many people as larger school districts. We wanted to do some type of an end-of-the-school-year theme, and all teachers had to teach about that theme. Like, this year we wanted to do the Olympics, but we just didn't get to it.
>
> Participant 3: Yeah, we've talked about wanting to do a thematic unit and having like the theme be ancient Greece or the Olympics or something. And then trying to incorporate all of the different subjects into that. We didn't get it done this year, but ***we did a lot of talking about it***.

Faculty as work context. Not surprisingly, teachers described their faculty peers as important factors in shaping the character of their workplace. However, several issues in these rural schools stood out. First, while teachers sought out their colleagues to address questions of practice, these interactions were often spontaneous and unplanned and did not focus on deep examinations of teachers' behaviors and beliefs. Teacher interactions typically addressed mundane matters and with minimal impact on teacher practice. This is not to say that teachers were irresponsible or intransigent; it only suggests that learning aimed at supporting changes in practice did not typically occur via faculty interaction.

Contrary to the image of faculty cultures as deeply rooted constructs, teachers at each of the schools described cultures that were somewhat unpredictable. Teachers argued that the small faculty size and isolation of the schools led to a revolving-door phenomenon among faculty that could easily destabilize school cultures. For example, a faculty culture could become transformed over one summer with a few key faculty retirements. Several teachers described this as a benefit of working in a small rural school because it allowed the school to hire teachers with fresh ideas.

On the other hand, from the perspective of teachers who had decided to make a school their professional home, the flow of faculty in and out of the school created a challenging environment. In two of the schools, several teachers described how their departmental colleagues were all novices. The veteran teachers found that while discussions with their more novice peers were often engaging, they found themselves more often than not in mentoring roles. While they accepted this role and enjoyed mentoring new faculty, their learning was not of the same depth and reflection as might have occurred with teachers of their same experience level. As one teacher put it:

> It's me and then two first-year teachers in the language arts department. So there isn't a whole lot of collaboration because they're so new that they're just working on getting through....Before we had two, more experienced teachers....We did more talking with each other and suggesting or asking, what kind of novels would you suggest for this?

Ironically, while the small size of these rural schools facilitated communication among all teachers, teachers still operated within limited spheres of interaction. In other words, even in schools with faculties of 18 teachers, teachers still worked primarily with a small number of colleagues. Within these spheres of interaction, teachers described norms that guided the course of relationships and often bounded the conversations. For example, most teachers' collegial relationships focused more on discussions around classroom logistics, curriculum, even content, and less on their successes or failures with particular students and their approaches to teaching. Where discussions among teachers were focused on teaching, they swirled around the surface and did not penetrate deeply into the core context. For example, teachers shared materials, provided interdisciplinary support (e.g., art and English), but rarely described in-depth sharing or challenging assumptions in each other's practice. One teacher's comments capture this hesitancy to pry too much into the practice of other teachers:

> Our high school has very good working relationships. We're all on different committees or we're assigned activities together, and we all work pretty well together. As always, you'll have some teachers that you don't feel do their job correctly, but you don't want to say anything to anybody. They'll figure it out.

Teachers also described how the reluctance to engage or even challenge peers in substantive ways affected faculty relations in negative ways. For example, at one school several teachers mentioned the damage to faculty motivation that one incompetent teacher brought. That incompetence caused a ripple effect that was felt within the core context. As a result of one teacher's failures, another teacher's students in the same subject area were influenced each year. The teacher's focus was on remediation and not the teaching of the material she was meant to teach. Further, her motivation to learn more deeply the subject matter she was supposed to be covering was lacking. She said:

> I had students in Spanish 4 when I first got here that could not even complete an entire sentence, could not respond to "como estas," for example. And so that was very frustrating for me because I came in and I knew how long she had been here. I knew she had her masters. She was making twice as much as me, and she didn't do anything while she was here for three years. So that is the kind of thing that I find frustrating. For me, that is one thing that will make or break whether I stay. I'm very self-motivated. I keep going. I really want my kids to learn, but if you constantly see people getting rewarded for not doing what they're supposed to be doing, it impinges on what you're trying to do.

However, while the competence level of co-workers was an important characteristic of work context, teachers did little on their own to confront faculty unable to meet the needs of students. Incompetence, while seen, was not discussed. "Intrusions" into the practices of peers were acceptable only when invited or through some formal channel, such as a mentoring program.

The Principal as Context. This study showed that principals as part of the intermediate context can influence teacher work and learning in several ways. The principal's engagement in the classroom and focus on teachers' concerns enhanced teacher learning within the core and intermediate contexts. It was clear that the small size of these rural schools (and consequently the small number of faculty) increased the potential for principal-teacher interaction, and certainly provided opportunities for principals to support and foster teacher learning. Through their actions or inactions, principals influenced the students-as-context, teachers' perceptions of subject matter, and teachers' relationships with their colleagues.

While principal activism does not guarantee a positive influence on work context, in this study the one principal who took an active role did create learning opportunities for teachers. The principal's positive influence on school culture was compared to the previous principal who had become overly passive at the end of his career. Under the previous principal, the school culture had turned dysfunctional as teachers took an "us versus them" stance toward their students; and, according to teachers, students disregarded basic rules of conduct by fighting in hallways, disrespecting teachers, and coming and going at will during the school day. Juxtaposed against this scenario, the new principal was a "change agent" who immediately influenced school culture. First, this principal influenced student-as- context in ways that refocused teachers' perspectives of students in much more constructive ways.

The new principal began to institute policies that required students to be responsible and accountable for their actions. By gaining control and reorienting the student culture, teachers' described a shift in their mindsets that had been teacher-centered to more student-centered approaches to teaching. Furthermore, teachers began to feel empowered in the school. As one teacher said, "Finally, we're running the school with input from the students." Teachers attributed this positive change to the principal. One teacher explained:

> He's got me seeing the possibilities for each student, even for some of them that are squirreling around. Why would a kid who's squirreling around that morning in first hour class—why did they come in at the end of the day? Well, it's because they know they can, and it's a safe place. So you can grab him and say, you know, you're being a real dork. Why don't you get to work. And he'll go with that; he can accept that. So I guess [the principal's] whole message...he's not real hung up on how much English did you teach today in 10 minutes?....It's more of how did you get into their heads?...I appreciate the fact that he has said that to us, that our curriculum or our subject matter is how we get into these kids' heads.

Second, compared to his predecessor, this principal blurred the boundaries between the roles of teachers and of the principal. Teachers described the rejuvenating effect of a principal who would spontaneously teach side by side with teachers. One teacher described it this way, "He'll just walk in the room and he and I just have this way we can just play off each other, which is very exciting because we did a little team teaching." In a later interview, she added:

> I did this survey with these kids...it was mid-year which I was really seeing some attitudes and hearing about some behaviors and stuff that were really disturbing to me....I was very pleased with what I found; they just want to be thought of as very good, and they want to do what is right, but their actions don't show that. I took the results to [the principal]...and boy that just got his brain working. We ended up team teaching our seniors for three days together.

Initially a shock, teachers soon began to welcome the principal into their classrooms as a colleague. These interactions and the principal's breadth of subject matter (especially math and science) and pedagogical knowledge created avenues for the principal to motivate and empower teachers to learn about their own subject matter. Teachers argued that the expectation of unexpected classroom visits had a positive effect on teachers because they had grown complacent and teacher-centered under the previous principal.

Finally, the principal played a critical role in shaping the ways teachers interacted with each other. Through his actions in the classrooms and efforts to act as a conduit between teachers, the principal was able to begin to develop a foundation upon which faculty could build their own relationships. For example, one teacher described the value of an off-campus retreat during summer break that the principal sponsored to help the staff get to know each other and "how they clicked." Also, as a result of his constant presence in classrooms, the principal was able to help teachers see connections and possibilities among them that were previously overlooked. Through means such as these the principal communicated to the faculty his belief in the importance of investing in and developing strong professional relations among the staff. He also modeled that fact by engaging groups of teachers in informal conversations focused on student issues, content, and pedagogy. In short, this principal exhibited what Stein and D'Amico (2000) identified as an aptitude for subject matter leadership. He was an excellent teacher and could translate his skills and knowledge base about teaching for teachers in any subject area. For example, compared to the other high schools studied, teachers at this school were more involved in learning with each other, the norms of secrecy were breaking down as the principal made his presence known in each classroom. Beyond presence, however, the manner in which the principal spontaneously engaged himself into the flow of the classroom helped shaped teachers' perspectives of themselves as learners. Put differently, the principal acted as "principal colleague" for teachers, many of whom had no (or few) other colleagues in their subject areas.

On the other hand, in the other two schools teachers perceived the principals as less activist. A principal's perceived inaction vis a vis teachers' work context is neither good nor bad, necessarily. However, in two of the case study schools principals' inaction with regard to developing contexts that supported vigorous teacher learning climates may have reinforced traditional approaches to teacher learning. For example, in the case regarding an incompetent teacher, the principal's and district's lack of disciplinary action lowered staff morale, and increased the need of another teacher who taught the upper level sections of that subject to remediate students. Unfortunately this situation lowered her enthusiasm for pursuing her own learning. In the third school, the principal played a minimal role in facilitating teacher learning by limiting his activities to authorizing travel and reimbursing participation in formal professional development activities. In this case the principal and teachers tended to see the principal's role as traditional—e.g., professional development through formal evaluation.

Peripheral Context: The District and State Contexts

It was apparent that districts and the state policy milieu also formed facets of teachers work context. Both the districts within which teachers worked and the policy milieu—that is, state policy initiatives that were designed and able to influence teacher work—formed part of teachers' learning contexts, but not necessarily with the immediacy of the core and intermediate contexts.

The district-as-context influenced teacher learning in two ways. First, teachers across schools believed that district leadership was responsible for setting the agenda for formal professional development. In describing the districts' role in professional development, the tension teachers felt between meeting individual teacher needs and organizational needs was palpable. A second factor influencing teacher learning at the district level were professional development committees (PDCs) that consisted solely of teachers. State law required that each district fund professional development to at least one percent of the basic district funding formula and that each district establish a professional development committee responsible for allocating those funds.

Teachers at each of the districts experienced similar patterns regarding the organization of their learning. Teachers identified several problems. In each case, districts had allotted approximately 5 professional days throughout the school year. In order to stretch resources and achieve economies of scale, the districts often held district-wide professional development events. While some teachers benefited from the workshops and speakers, most teachers interviewed found the events to be ineffective. For example, because most teachers in the district represented primary and middle school grades, teachers in this study often found the activities to be irrelevant to their needs. These teachers believed that by casting a wide net to accommodate all teachers, learning became superficial. Generally, teachers were frustrated by the waste of time spent at district professional development activities. A language arts teacher's comments capture this tension:

> I dread the PDC days, and I've been saying that since last year. I get in trouble with elementary teachers. I don't think they're crazy about me. But I don't know why we all meet as a group. I do think we're in this together, and I think every now and then we do need to meet as a group and say, what are you teaching, what are we doing, where can we go from there?

One of the high school's PDC representatives described the tension this way:

> The high school people are outnumbered since we have three elementaries. The reading workshop's kind of good, because like I said the elementary people felt it was useful mostly, and they are 3/4 of our population of teachers, and our high school people were sitting there kind of like....And if we had turned that around and had something that was more specific for high school that didn't apply to the elementaries, they would have been equally rude or disgruntled. So it is difficult to come up with something that works for the whole district.

By and large, however, teachers approved of districts' PDCs. PDCs were valued by teachers in a few ways. For example, teachers believed that the districts' professional development committees (PDC) provided the freedom for teachers to pursue their own learning goals even if those goals were not completely aligned with district goals for teacher learning. Teachers explained that their interests in attending regional, state, or local conferences were usually supported by the PDC if their requests were within the committee's budgetary constraints. Furthermore, the types of teacher learning activities funded by the PDCs offered teachers the opportunity to "benchmark" their own practices against their peers' across the state and nationally. Benchmarking practice was a theme among teachers across schools that highlighted the important role of oft-maligned large-scale conferences. The venues for learning served as a way to overcome the isolation of professional practice in these small rural schools.

From teachers' perspectives, the predominant state-level feature influencing teacher work was the state's program for performance assessment. The purpose of the assessment program was to increase student achievement levels by increasing the levels of school and district accountability. While the intended effects on schools and teachers were to change multiple aspects of professional practice (e.g., pedagogy, content knowledge, and assessment practices), interview data suggested that the density of the core contexts often deflected these intentions. Furthermore, the program's influence was limited by its focus on teachers of the core academic subjects. As a result, it was not surprising that core subject teachers discussed the assessment program and its influence on them more than teachers of career and technical subject areas. However, teachers' descriptions of how their practices had changed as a result of the program were telling because the changes do not penetrate deeply into multiple aspects of the core context. As interview data repeatedly showed, teachers reacted to the new performance assessment primarily by changing the way they assessed students in some cases. A language arts teacher described a typical response to the state's assessment program:

[The assessment program] is about critical thinking and comprehension….It's altered the way that I approach mostly my testing. I try to [reflect the assessment program] in my test questions even though my grade levels do not take it in communication arts. I've tried to support that. We've had training. We've had an [assessment program] leader here, and we will next year also.

This tweaking on the margins of teacher practice did not appear to influence teachers' thinking of their subject area, content knowledge, or student needs for leading productive lives in the future. Instead, with observations of fairly traditional teacher directed learning environments, most teachers articulated a presumption of compliance with the policy intent of the program. That is, teachers believed that they already taught critical thinking skills and changed their approaches to their students in preparing them for taking the high stakes tests. For instance, an English teacher stated:

> I'm fresh out of college, so I know what skills they need to go in there. So I look at real life, what are they going to need, and teach to that, and thankfully that's what's on the [State Assessment Program (SAP)]. We do the whole writing process, the revising, webbing, graphic organizers, and those are very important on the SAP....If a teacher is truly doing his or her job, then they are naturally teaching to the test. What I don't do is, I'm not going to review for five weeks and think, "I've covered what's on there."

In part, the muted impact of the policy milieu on the classroom context may stem from teachers beliefs that their students needs are much broader than academic. Teachers argued that the assessment program was too distant from the classroom context, and they believed that district and school policies were better situated to keep the students' best interests in mind. As one teacher stated:

> I would rather see student achievement done on a portfolio-type basis, working out of the school district, not so much a test to see where we stack up against other [districts]. I would rather see more local accountability. More local processing. I feel a school district knows its children the best, better than the state can know them...And I really feel it needs to be done more on a tracking of growth through a portfolio, maybe that they begin in seventh grade and keep going up through twelfth grade and look at that as an evaluative instrument rather than a test they take two or three mornings.

Discussion

This study contributes to the literature by taking the concept of teacher work context—a concept that has become synonymous with structural and organizational dimensions of schools—and recasting it in terms of sets of relationships. Furthermore, by examining these relationship sets within the context of rural high school teachers, several assumptions about teacher learning became apparent. For example, from teachers' perspectives the gap between teacher and student knowledge is fluid and not confined to student academic performance. Also, while the virtues of small schools and small professional communities are extolled in the literature, the smallness and isolation of the rural context created significant challenges to teacher learning and elicited specific teacher learning responses. Finally, current discussions of teacher learning tend to lightly address, if at all, the critical role of the principal and other educational leaders in supporting teacher learning, especially in rural schools where subject matter colleagues are limited in number.

This article showed how the multiple contexts of teacher work and their relationship to teacher learning are complex and interwoven. To state the obvious, each school context is unique. For instance, in these rural high schools the department was not the context of most import as suggest in other studies (e.g. McLaughlin, 1993). Most teacher did not have departments, per se. The context of most import for them was the classroom, or at times, their professional community external to the school and district. Furthermore, this study shows the complex relationship between student-teacher and teacher-principal relationships in shaping teachers individual and collective beliefs about their students.

Conceptualizing teacher work context as beginning with a dense core and emanating outward to levels of context provides an important image that is less prone to over-simplifying how teachers do and should learn on the job. More than any other factor, the "hot action" and "here-and-now" urgency of the classroom life (Eraut, 1994; Jackson, 1968) stemmed from the complex and daily interactions of students. These interactions, marked by the fluidity and unpredictability of student needs, attitudes, and behaviors, are what give the core context its density. And it is this density that causes efforts to support teacher learning from the intermediate and peripheral contexts to be met with resistance. Teachers' subject matter served as the conduit through which most teacher-student interactions took place. In this study, teachers of subjects with well-defined knowledge bases such as math and science also tended to define the triadic relationship of the core context narrowly. Their efforts were focused primarily on transferring the conceptual knowledge. Other teachers—e.g., language and social studies teachers—used the content of their subject area to broadly address student needs. For example, these teachers would use the content to explore ethical dilemmas in students' lives as well as to convey the specific content.

The intensity of teacher-student interaction and the way teachers perceived of their subject matter created an environment in which teachers believed that only the most practical and immediately applicable knowledge/skill would suffice to inform their practice. Thus, because of the intensity of interaction between teachers, students, and content, learning that was valued most was acquired experientially.

The intermediate context played an important role in shaping the broader context of teachers' everyday lives. The principal was uniquely situated to influence important relationships within the core and intermediate levels. For example, the principal could influence the core context by influencing student culture, shaping how teachers perceived students, shaping how teachers approached their own content, and influencing the nature of teacher relations. The principal also was uniquely situated to reshape the norms of professional practice as that practice related to students, subject matter and school culture, and in so doing influence the content and nature of teacher conversations about their work and students.

In spite of knowing what strong learning communities look like, however, traditional approaches to teacher learning persist. Contributing to this persistence were teacher attitudes and norms that hold back professional critique and shy away from opportunities to "deprivatize" practice (Louis, Kruse, & Marks, 1996). However, the blame cannot be solely placed on teachers. The important roles of principals were clear in this study. For example, the principal who broke down the metaphorical walls that isolated teachers was beginning to see more collaboration and learning among teachers. In schools where the principals took on passive leadership roles, teachers continued to work in (and worked to maintain) isolation. Furthermore, teachers emphasized how resources strain limited the possibilities of periodic interaction among peers.

In short, the potential of intermediate level factors for influencing the core context was significant. The potential was found in the possibilities of school leaders and teachers' peers in shaping how teachers perceived of their students, how they taught their students, and the way they approached their subject matter in terms of their students. Furthermore, in small and isolated high schools such as these where collegiality within departments was often a moot issue, the role of the principal as "principal colleague" and facilitator of teacher collegiality was critical to the health of professional learning.

The peripheral context of teacher work did not factor into teachers' work lives on a daily basis. Nonetheless, the study uncovered several findings that cast new light on teacher learning in rural schools. Most surprising was that in districts where central offices were often located on the same campus as the high school, teachers perceived of the superintendent and district as peripheral to their learning. Without a doubt, district policies surrounding professional development influenced teacher work context. Unfortunately, district sponsored professional development programs were of little relevance to high school teachers. On the other hand, teachers found that state legislated district support for individual learning was more useful. Thus, district professional development committees acted as vehicles for teachers to broaden their professional networks by attending regional and state conferences. This fact underscored a difference between these rural teachers and urban and suburban teachers (Scribner, 1999; Talbert & McLaughlin, 1996). That is, teachers in these rural high schools relied to a greater extent on professional community beyond the school walls. Because departments in these schools were small, opportunities for fruitful collegial relations were often found elsewhere.

Statewide performance assessment was the major state policy initiative that influenced many teachers in this study. However, by definition, only teachers in the core academic subjects were influenced by the program. In some cases this policy influenced the core context of teacher work. However, the effect was marginal in that teachers described minor changes to their practice—i.e., changing the way they assess students. Many of the teachers interviewed were not in subjects that were assessed and, obviously, they were not influenced by the policy.

Thinking about the relationship between teacher learning and work context in terms of three concentric contexts brings into focus several important implications for teacher learning. Teacher learning within the core context is inevitable. The question remains, though, how will that learning occur and what will it focus on? Left alone, teachers have no choice but to learn reactively, guided only by their own experience and the immediate needs of students. Somewhat paradoxically, professional development for principals, superintendents and other educational leaders may have important implications for teacher learning. While professional learning opportunities for teachers abound, opportunities for principals and superintendents occur with less frequency. Professional development for educational leaders must place more emphasis on their roles in shaping the learning cultures of schools and more specifically subject matter leadership (Stein & D'Amico, 2000). Furthermore, opportunities for teacher professional development must consider the situations of rural teachers. For teachers in this study, statewide professional conferences were places for establishing and maintaining professional community. And finally, the development of virtual professional development opportunities may be effective strategies for bringing teachers of similar subject matters into professional conversations.

In summary, this study demonstrated how the density of the core context was defined by intense relationships among students, teachers, and content. For these relationships to be nurtured and shaped in ways that improve teacher practice and student learning they must be (1) acknowledged in the broader education context, (2) understood as a strength in the teaching and learning process, and (3) addressed in a concerted way across all three contexts. If teachers are left to "figure things out" on their own in isolation, most of them will be forced to learn solely through experience and in unreflective ways. The role of the intermediate and peripheral contexts in teacher learning must be to provide the time, space, and support for practical and conceptual learning that teachers believe (or come to believe) is meaningful within their more immediate work context—the relationship between teacher, student and subject matter.

Acknowledgement

This article was supported entirely through the National Academy of Education/Spencer Foundation Post-doctoral Fellowship Program. An earlier draft was presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 1-5, 2002.

References

Bredeson, P. V., & Scribner, J. P. (2000). A statewide professional development conference: Useful strategy for learning or inefficient use of resources. *Educational Policy Analysis Archives, 8*(13). Retrieved on March 25, 2003 from http://epaa.asu.edu/epaa/v8n13.html.

Denscombe, M. (1980). Keepin' em quiet: The significance of noise for the practical activity of teaching.

Denscombe, M. (1982). The hidden pedagogy and its implications for teacher training. *British Journal of Sociology in Education, 3*, 249-265.

Eraut, M. (1994). *Developing professional knowledge and competence*. London: Falmer Press.

Fullan, M. G. (1995). The limits and the potential of professional development. In M. Huberman (Ed.), *Professional development in education: New paradigms and practices* (pp. 253-267). New York: Teachers College Press.

Grant, C. A., & Sleeter, C. E. (1987). Who determines teacher work? The debate continues. *Teaching & Teacher Education, 3*(1), 61-64.

Hargreaves, A., & Goodson, I. F. (Eds.). (1996). *Teachers' professional lives*. London: Falmer Press.

Hatton, E. (1987). Determinants of teacher work: Some causal complications. *Teaching and Teacher Education, 3*(1), 55-60.

Hawley, W. D., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In G. Sykes (Ed.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 127-150). San Francisco: Jossey-Bass.

Huberman, M. (1993). The model of the independent artisan in teachers' professional relations. In M. W. McLaughlin (Ed.), *Teachers' work: Individuals, colleagues, and contexts* (pp. 11-50). New York: Teachers College Press.

Jackson, P. W. (1968). *Life in classrooms*. New York: Holt, Reinhart & Winston, Inc.

Jackson, P. W. (1986). *The practice of teaching*. New York: Teachers College Press.

Johnson, S. M. (1990). *Teachers at work: Achieving success in our schools*. New York: Basic Books.

Little, J. W. (1999). Organizing schools for teacher learning. In G. Sykes (Ed.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 233-262). San Francisco: Jossey-Bass.

Louis, K. S., Kruse, S. D., & Marks, H. M. (1996). Schoolwide professional community. In F. M. A. Newmann (Ed.), *Authentic Achievement: Restructuring schools for intellectual quality*. San Francisco: Jossey Bass.

McLaughlin, M. W. (1993). What matters most in teachers workplace context? In M. W. McLaughlin (Ed.), *Teachers' work: Individuals, colleagues, and contexts* (pp. 79-103). New York: Teachers College Press.

Scribner, J. P. (1999). Professional development: Untangling the influence of work context on teacher learning. *Educational Administration Quarterly, 35*(1), 238-266.

Scribner, J. P., Hager, D., & Madrone, T. R. (in press). The paradox of professional community: A tale of two high schools. *Educational Administration Quarterly*.

Stein, M. K., & D'Amico, L. (2000). How subjects matter in school leadership. Paper presented at the annual conference of the American Educational Research Conference: Seattle, WA.

Sykes, G. (1999). Teacher and student learning: Strengthening their connection. In G. Sykes (Ed.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 151-179). San Francisco: Jossey-Bass.

Talbert, J. E., & McLaughlin, M. W. (1996). Teacher professionalism in local school contexts. In I. F. Goodson (Ed.), *Teachers' professional lives*. London: Falmer Press.

Thompson, C. L., & Zeuli, J. S. (1999). The frame and tapestry: Standards-based reform and professional development. In G. Sykes (Ed.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 341-375). San Francisco: Jossey-Bass.

Yin, R. K. (1994). *Case study research: Design and methods*. Thousand Oaks, CA: SAGE.

About the Author

Jay Parades Scribner
University of Missouri-Columbia
Email: scribnerj@missouri.edu

Jay Paredes Scribner is an Assistant Professor in the Department of Educational Leadership and Policy Analysis at the University of Missouri-Columbia. His research focuses on professional learning in K-12 contexts. Specifically, his research has focused on the relationship between teacher learning and teacher work and the political dimensions of professional communities in schools. He has also conducted research on professional learning in principal preparation programs. His other professional experiences include teaching in the Peace Corps, program evaluation and policy analysis with the US General Accounting Office, and evaluation consulting in the K-12 and higher education arenas.

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México

bracho dis1 cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

This article has been retrieved 5718 times since April 22, 2003

Education Policy Analysis Archives

Volume 11 Number 13 **April 22, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright 2003, the **EDUCATION POLICY ANALYSIS ARCHIVES**. Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Massachusetts Signing Bonus Program for New Teachers: A Model of Teacher Preparation Worth Copying?

R. Clarke Fowler
Salem State College

Citation: Fowler, R.C. (April 22, 2003). The Massachusetts Signing Bonus Program for New Teachers: A model of teacher preparation worth copying? *Education Policy Analysis Archives, 11*(13). Retrieved [date] from http://epaa.asu.edu/epaa/v11n13/.

Abstract

This article examines the Massachusetts Signing Bonus Program for New Teachers, a nationally prominent program that has recruited and prepared $20,000 bonus recipients to teach after seven weeks' training at the Massachusetts Institute for New Teachers (MINT). Although state officials have trumpeted this initiative as a national model that other states are copying, they announced in November 2002 that they were radically changing it. The changes included halting the state's national recruitment efforts and replacing the seven-week, fast-track training program designed by the New Teacher Project with year-long programs to be designed by three of the state's education schools. Even though the state spent more than $50,000 recruiting individuals from states outside the Northeast over the first four program years, it garnered just seven bonus recipients from the non-Northeast states its recruiters visited, only four of whom were still teaching in Fall 2002. The state did, however, generate a substantial number of applicants in each program year (ranging from 783 to nearly 950), most of whom came from Massachusetts or nearby states. Contrary to state officials' claims, though, it appears that many of these individuals had substantial prior educational experience. Although officials stated that all bonus teachers would go to 13 designated high-need urban districts, the state has never met this commitment, sending fewer bonus teachers to these districts in each of the first three years of the program. The state has lost a high percentage of its bonus teachers to attrition particularly in state-designated, high-need districts. These attrition rates are substantially higher than comparable national rates. Although the state has portrayed the Bonus and MINT programs, combined, as highly successful, officials exaggerated many of the purported positive outcomes. On the positive side, independent survey data (Churchill et al., 2002) indicated that principals rated MINT graduates' performance favorably, when compared to traditionally-trained teachers. It is not clear, though, whether such ratings varied either by a) the extent of the teacher's prior educational experience or b) the nature of the teacher's placement (urban vs. suburban). The Bonus Program has produced relatively few urban teachers, relatively few minority teachers, and low rates of teacher retention, even though this effort was modeled after Teach for America and critical parts of it were designed and often managed by the New Teacher Project—two organizations that the Bush administration has praised for their ability to design and run programs of this type. Policy makers are urged to resist calls to embrace rapid certification, an approach that has produced, in Massachusetts, low numbers of urban teachers and high numbers of exiting teachers, all at a cost of more than four million dollars.

Improving teacher supply and quality is among the most important educational issues the United States faces today as policy makers seek to hire, over the next ten years, both more and better teachers than ever before (Zernike, 2000). In Massachusetts, state officials have developed a number of initiatives to address this goal, the most notable being the Massachusetts Signing Bonus Program for New Teachers. This fast-track alternative to traditional certification programs has garnered national attention for its $20,000 bonuses to selected candidates, its cross-country recruiting campaign, its intensive seven-week training provided by the Massachusetts Institute for New Teachers (MINT), and its claims of success (School Board News, 2001).

But the Bonus Program has also generated significant controversy. In February 2001, independent research suggested that the Bonus Program had failed to live up to policy makers' promises. Twenty percent of its first cohort of bonus recipients left teaching after one year, and less than half of its second cohort chose to teach where policy makers said they would—in 13 state-designated, high-need school districts (Fowler, 2001).

The Massachusetts Commissioner of Education, David Driscoll, steadfastly defended the Bonus Program against the preceding criticisms, claiming that this initiative was "hugely successful" (Setera, 2001). In November 2002, however, Commissioner Driscoll announced radical changes in the Bonus Program. Starting in March 2003, bonus recipients will be both nominated and trained by three "innovative" year-long post-baccalaureate teacher preparation programs (Massachusetts Department of Education, 2002a). The state will continue to run the MINT program, but Bonus recipients will not attend it.

With these changes, the Bonus Program silently dropped its cross-country recruiting campaign, stopped participating in the MINT training program, terminated its relationship with the New Teacher Project (the organization that designed and managed many aspects of this initiative), and abandoned the 7-week, fast-track approach to training bonus recipients. Massachusetts has now turned to the state's schools of education, formerly portrayed by the state's policy makers as the source of poor teacher quality, as the sole source of new teachers worthy of $20,000 bonuses.

Why did one of the first states to jump with both feet into fast-track certification pull one foot out? According to Orin Gutlearner, who coordinates the Bonus and MINT programs, this change represents "a better use of our resources" (Archer, 2002). Commissioner Driscoll stated that "by making this change in our program ... we will not only be able to support our colleges and universities, but also be better equipped to truly find the most qualified recipients for our signing bonus" (Massachusetts Department of Education, 2002a).

The preceding statements fail to answer precisely why the state made radical changes to its famed bonus program. The current paper investigates, however, factors that surely influenced state policy makers' decision to radically alter this program. The data discussed here were obtained from the Massachusetts Department of Education, either through its web site or through a series of Freedom of Information requests submitted to the Department between 1999 and 2002.

Two factors warrant scrutiny of the Bonus Program. First, the Commonwealth of Massachusetts has invested a substantial amount of money, more than $4 million, for this initiative that policy makers have used, in part, as a template for recruiting and training at least ten percent of the state's new teachers (Commonwealth of Massachusetts, 2001). It is critical to gather information that will allow Massachusetts legislators and the public to assess the effectiveness of this major effort.

Second, the federal government is vigorously promoting fast-track alternative certification programs that, like Massachusetts's Bonus Program, "recruit highly qualified candidates who are interested in teaching but did not attend schools of education and place them quickly into high-need schools, providing training, support and mentoring ... These programs should become models for the future, as states make it less burdensome for exceptional candidates to find teaching positions in our nation's schools" (U. S. Department of Education 2002, p. viii). Indeed, the federal government is endorsing not just the fast-track approach implemented in Massachusetts, but the organization that inspired this initiative, Teach for America, and its offshoot, the New Teacher Project, that the Massachusetts Department of Education hired to plan and implement central elements of the Bonus Program.

For the Bush administration, the fast-track approach to teacher preparation constitutes a "more promising model for the future" (U. S. Department of Education 2002, p. 2), a model that, if adopted nationally, will lead to the production of both more and better teachers and thereby make it possible for the United States to place highly qualified teachers in all schools, particularly challenging urban schools. It is important to investigate, therefore, what this approach has delivered to Massachusetts and, concomitantly, what it promises to deliver to the country.

Background

The Massachusetts Signing Bonus Program for New Teachers was introduced in the summer of 1998 in the wake of the great uproar over the results of the first two administrations of the Massachusetts Teacher Tests (now called the Massachusetts Tests for Educator Licensure). Concerned both about the quality of the state's teaching force and about looming shortages of teachers in particular content areas (such as math and science), the legislature and then governor Paul Cellucci supported legislation that established a $60 million endowment to improve teacher quality. The most prominent feature of this legislation was a fast-track certification program that was initially called Teach for Massachusetts (Massachusetts Department of Education, 1998) but was later named the Massachusetts Signing Bonus Program for New Teachers.

The stated aim of the Bonus Program is "to encourage high achieving candidates to enter the profession who would otherwise not consider a career in teaching" (Massachusetts General Laws, Chapter 15A, Section 19A). The form of encouragement the state offers is a $20,000 bonus signing bonus, with $8,000 distributed in the first year and $4000 in each of the following three years.

The Massachusetts Department of Education contracted with the New York-located New Teacher Project, an offshoot of Teach for America, to design and oversee the recruitment and selection of the first cohort of bonus teachers. The Department named the recipients of the first round of bonuses in spring 1999 and trained 59 individuals in an intensive seven-week summer training program at the Massachusetts Institute for New Teachers (MINT). These individuals began teaching in Massachusetts public schools in fall 1999.

In order to increase the number of MINT-trained teachers in the following years, the Department 1) increased the number of bonus recipients to over 100, 2) offered scholarships (worth $2,250) to selected applicants who applied for but did not receive a bonus, 3) encouraged school districts to sponsor individuals' MINT training, and 4) allowed individuals to attend MINT at their own expense. The state has now recruited and trained four cohorts of bonus recipients (FY 1999-2002), as well as three cohorts of scholarship recipients and other MINT graduates (FY 2000-2002).

Organizations that either advocate for or are involved in fast-track teacher certification have lauded Massachusetts's Bonus Program. The Milken Foundation recognized it as a model for other states to follow (Hayward, 2000). The National Council on Teacher Quality (2002) listed Massachusetts as a "place to watch" for its recruiting practices. The New Teacher Project, which took the lead in conceptualizing and implementing all aspects of the state's effort, listed its work with Massachusetts as the first of four project highlights on its web page (New Teacher Project, 2002).

In February 2001, the Massachusetts Commissioner of Education, David Driscoll, stated that many other states have followed the Massachusetts model: "The immense influence of the potential of our new teacher recruitment program has influenced California, Connecticut, Florida, Maryland, New York and other states to initiate similar programs. We are encouraged by the fact that other states are copying what we are doing and have judged that what we are doing is worth copying" (Massachusetts Department of Education, 2001a).

But should other states copy the recruitment and training methods that Massachusetts has employed in its Bonus and MINT programs? Did the Bonus Program merit the praise it garnered? What factors influenced state officials to radically alter this program? To answer these and other questions, I consider next the Bonus Program's record in a number of crucial areas: recruitment, placement, attrition, and program effectiveness.

Recruitment

One element of the Bonus Program that garnered much national attention was its recruiting campaign, particularly its cross-country recruitment efforts. Headlines appeared across the country describing how Massachusetts was scouring the country, looking for "the best and the brightest" to come teach in Massachusetts. Although a number of articles indicated that this effort was highly successful (Ferdinand, 1999; Magee, 1999), data obtained from the Massachusetts Department of Education indicate that Massachusetts's efforts to recruit bonus recipients from states outside the Northeast (i.e., outside New England, New York, and Pennsylvania) have been ineffective. As Table 1 indicates, through the first four years of the program (FY 1999-2002), the state garnered just seven recruits from non-Northeast states, despite recruiting in five non-Northeast states in 1999, seven in 2000, nine in 2001, and four in 2002—at a cost of more than $50,000. Since three of these seven recruits have now stopped teaching, Massachusetts has spent more than $12,000 just to recruit each of the four active bonus teachers from these states.

Table 1
Cost of Bonus Teacher Recruitment in Non-Northeast States

Fiscal Year	Costs			Number of states visited	Number of recruits	Number currently teaching	Cost per current teacher
	Travel	Personnel (Note 1)	Total				
1999	$7,571	$3,685	$11,256	5	3	1	$11,256
2000	$8,865	$4,656	$13,521	7	2	1	$13,521
2001	$14,513	$2,631	$17,144	9	2	2	$8,572
2002	$6,943	$1,240	$7,943	5	0	0	-
Total	$37,892	$12,960	$50,852	21	7	4	$12,713

Ironically, as described in Table 2, between 1999 and 2002, Massachusetts actually recruited more than twice as many bonus teachers from non-Northeast states where they did not send recruiters than from states where they did (18 vs. 7).

Table 2
Number of Bonus Recruits from Non-Northeast States

Fiscal Year	Number of recruits from non-Northeast states where Massachusetts:	
	Sent recruiters	Did not send recruiters
FY 1999	3	3
FY 2000	2	6
FY 2001	2	5
FY 2002	0	4
Total	7	18

Why did Massachusetts experience such difficulty with its out-of-state recruiting? One probable factor, discussed in a prior study (Fowler, 2001), is that many of the areas where Massachusetts sent recruiters are experiencing teacher shortages that are far more severe than in Massachusetts. These states include Texas, California, and Florida.

Another factor may be that Massachusetts' recruiters frequently visited states where the average starting salary, when adjusted for the cost of living, is substantially higher than in Massachusetts. According to Education Week (2002), Massachusetts' average starting teacher salary, adjusted for the cost of living, is $26,565 and ranks 30th in the nation. Over the last four years (1999-2002), the state's recruiters have frequently visited higher paying states, including nine of the twelve top-paying states. New teachers in these nine states can expect to earn, on average, from $2,662 to $6,120 more per year than new teachers in Massachusetts.

Table 3 lists a) the states that Massachusetts's recruiters have visited, b) how Education Week (2002) ranked each state's starting salary on a national scale (adjusted for the cost of living), c) each state's average starting teacher salary, and d) the difference between the average starting teacher in that state and in Massachusetts.

Table 3
States Where Massachusetts Recruited for Bonus Teachers (1998-2002).

State	*Rank (1-51)*	*Average Starting Salary**	*Difference from Massachusetts*
Georgia	2	$32,685	$6,120+
Texas	3	$31,568	$5,003+
Pennsylvania	5	$30,911	$4,346+
Michigan	6	$30,878	$4,313+
Illinois	7	$30,745	$4,180+
North Carolina	8	$30,529	$3,964+
Delaware	10	$30,113	$3,548+
Indiana	11	$29,306	$2,741+
New York	12	$29,227	$2,662+
U.S. Average	--	*$27,989*	*$1,424+*
Virginia	20	$27,383	$818+
Wisconsin	21	$27,339	$774+
Washington, DC	26	$26,896	$331+
Florida	28	$26,631	$66+
Massachusetts	30	$26,565	$ -------
New Jersey	31	$26,542	$23-
California	36	$26,225	$340-
Rhode Island	39	$25,843	$722-

Connecticut	41	$25,352	$1,213-
Ohio	43	$25,017	$1,548-
Maine	46	$24,007	$2,558-
New Hampshire	47	$23,337	$3,228-

*Average starting salary is adjusted for the cost of living (Education Week, 2000).

A third factor that surely hindered the Bonus Program's national recruiting effort is the state's certification exam, the Massachusetts Tests for Educator Licensure (MTEL). While most states offer the Praxis, an exam that is administered throughout the United States, the MTEL is unique to Massachusetts. Therefore, when Bonus Program recruitment began in 1999, candidates had to travel to Massachusetts to take (and pass) the MTEL to be eligible to win a bonus.

State officials attempted to address this problem by offering the MTEL in seven cities outside Massachusetts on one date in FY 2001 and on two dates in FY 2002. As Table 4 indicates, however, this strategy met with limited success. The Department canceled the tests for three cities (Detroit, Houston and Miami) due to low enrollment. Tests were given, however, on each of the scheduled dates in Chicago, Los Angeles, Philadelphia, and Washington, DC. In all, 205 individuals sat for the MTEL in four cities in the last two program years (2001 and 2002).

Table 4
Number of Individuals Who Took the MTEL Outside Massachusetts.

Test site	Number who took the MTEL on:		
	January 2001	December 2001	February 2002
Chicago	16	10	14
Detroit	0 (Canceled)	0 (Canceled)	0 (Canceled)
Houston	0 (Canceled)	0 (Canceled)	0 (Canceled)
Los Angeles	11	21	22
Miami	0 (Canceled)	0 (Canceled)	0 (Canceled)
Philadelphia	24	18	20
Washington, DC	12	14	14
Total	63	72	70

Besides offering the MTEL in other states, the Department responded to its disappointing national recruiting campaign by cutting back on out-of-state trips in 2002, when they visited just four non-New England states (Georgia, North Carolina, Virginia, and Wisconsin) and Washington, DC. The Department continued to exercise questionable judgment, though, when deciding where to visit. Four of the preceding areas, unlike Massachusetts, produce fewer teachers than they themselves require. For example, North Carolina needs to hire 10,000 teachers, but expects to graduate just 2,200 from their education schools (Silberman, 2001); Virginia needs to hire 4,000 new teachers, but expects to graduate just 3,500 (Turner, 2001).

Data from Georgia illustrates the questionable decision to recruit teachers from that state. According to the Atlanta Journal Constitution, Georgia is desperate for teachers:

> Georgia schools hired 12,000 teachers this year, and the projected annual need is expected to reach nearly 20,000 by the end of the decade. The state's colleges of education only graduated about 3,500 students last year. Most hires come from out of state or are former teachers returning to the profession. (Salzer, 2002)

Moreover, Georgia's average teacher starting salary of $32,685 (adjusted for the cost of living) is $6,120 higher than in Massachusetts. Given this difference, an individual who chooses to teach in Georgia would effectively earn, over the next four years, $ 4,500 more than a Massachusetts bonus recipient would earn. Moreover, he or she would be able to enter the classroom more quickly in Georgia because that state's fast-track certification program, TAPP (Teacher Alternative Preparation Program), requires four weeks of training but Massachusetts' MINT requires seven.

Since Massachusetts's recruiters visited historically black colleges during their spring 2002 trips to Georgia, it appears that their aim was to recruit minority applicants. Although this is a worthy goal, only four individuals attended the Department's two recruiting sessions in Georgia (according to recruiters' notes I obtained from the Department via a FOIA request), and no one from Georgia applied to the program in that year (Churchill et al., 2002).

Despite the apparent failure of the state's national recruiting campaign, the state has nevertheless generated a considerable number of applications in each of its four program years: 783 individuals applied in 1999 (Massachusetts Department of Education, 1999a), "nearly 950" in 2000 (Massachusetts Department of Education, 2000a), 905 in 2001, and 932 in 2002 (Churchill et al., 2002). Not surprisingly, given the preceding analysis, most of the applications came from Massachusetts or neighboring states: specifically, in 2001 and 2002, 84% of applicants came from Massachusetts, and 93% came from New England or New York (Churchill et al., 2002).

With the recent changes to the program, though, the state has abandoned its bonus teacher recruitment efforts. Henceforth, three "innovative" post-baccalaureate programs will nominate candidates for the bonus and the state will make the final selection.

Placement

Although the state's initial set of recruitment materials clearly stated that all bonus recipients would teach in 13 high-need districts, (Note 2) Massachusetts has never met this goal. 71% of the first of bonus recipients cohort (FY 1999) were placed in these districts and 48% of the second cohort (FY 2000). In the third program year (FY 2001), only 35% of the bonus recipients went to the high-need areas. Table 5 shows that, from 1999 to 2001, the percentage of bonus teachers in high-need districts has steadily declined while the percentage in high-need districts has more than doubled.

Table 5
Initial Bonus Teacher Placement, Fiscal Years 1999-2001

Fiscal Year	Number of bonus teachers	Bonus teachers initially placed in:			
		High-need areas		Non-high-need areas	
		Number	Percent	Number	Percent
1999	59	42	71%	17	29%
2000	105	45	43%	60	57%
2001	105	37	35%	68	65%

When teacher attrition and migration are taken into account, 41% of all active bonus recipients in the 2001-2002 academic year taught in the state's designated high-need areas. This means, of course, that 59% of all active bonus recipients taught in non-high-need areas.

Many of the bonus recipients have taught not just outside of the designated high-need areas, but in the state's most wealthy and academically successful districts. In 2001-2002, 37 bonus recipients taught in school districts that scored in the top 50 of the Boston Globe's (2001a, 2001b) rankings of districts according to their scores on the 2001 Massachusetts Comprehensive Assessment System (MCAS). More than half of these bonus teachers, 21 in all, taught in nine elite districts that scored in the top ten of the Globe's 2001 MCAS rankings. These districts are: Belmont (5 bonus recipients), Concord-Carlisle (2), Lexington (3), Lincoln-Sudbury (1), Needham (1), Newton (5), Wayland (2), Wellesley (1), and Weston (1). Two of these highly advantaged communities, Belmont and Newton, had more bonus teachers together (10) than eleven of the communities that the state designated as high-need areas, i.e., Brockton (5), Cambridge (4), Chelsea (7), Fall River (2), Holyoke (1), Lawrence (3), Lowell (6), Lynn (6), New Bedford (2), Springfield (6), and Worcester (6).

Why did the promise of a bonus not result in the placement of teachers in high-need districts? At least two factors are at work here. First, many bonus recipients are not interested in teaching in the state's high-need areas. According to notes I obtained from the Department (via a FOIA request), a Boston principal informed a MINT evaluator that, when he (the principal) asked for a list of (FY 2000) bonus recipients willing "to teach in a urban school, the list was only about 1/3 of bonus recipients."

The second factor is the inconsistent nature of the state's commitment to high-need districts. When the media reported in February 2001 that many of the FY 2000 bonus recipients were not teaching in the high-need districts, two top state officials offered conflicting responses. When a television interviewer asked Commissioner Driscoll why so many bonus recipients were not teaching in these districts (where he had said they would teach at a spring 1999 press conference), Mr. Driscoll replied, "What I said is that was our goal" (Setera, 2001). Deputy Commissioner of Education Alan Safran told the Boston Globe, however, that the state did not have such a goal: "it is 'bizarre' to think qualified teachers should teach only in 13 urban school districts" (Vaishnav, 2001, p. B1). Later he told another reporter that "there are at least 50 districts that we would consider high need" (Associated Press, 2001). Each of the preceding statements contradicts, however, prior statements from the Department, (Note 3) rendering the goals of the program ill-defined at best.

The Boston Herald reported in January 2002, however, that the state had reaffirmed its previously denied commitment to sending bonus teachers to urban, high-need areas. Orin Gutlearner, the Department's Director of Alternative Teacher Recruitment, stated that the Bonus Program was now seeking "people with experience in urban communities and a mindset to work where they are most needed" (Hayward, 2002). This intention was reflected in the wording of the state's (FY 2002) recruiting materials that were subsequently on the Department's web site: "Priority will be given to candidates who are committed to working in an under-resourced urban district (i.e., Boston, Brockton, Fall River, Lowell, Springfield, and Worcester)" (Massachusetts Department of Education, 2002).

Problems remained, however, with the state's new-found commitment to high-need schools districts. First, at the same time that the Department renewed its commitment to high-need areas in the statement cited above, it dropped seven districts from its list of such areas. Specifically, it dropped Cambridge, Chelsea, Framingham, Holyoke, Lawrence, Lynn, and New Bedford. Although it is understandable that the state would drop Framingham, a district with higher MCAS pass rates and lower poverty rates than the other high-need districts, from this list, it is incomprehensible that it would no longer designate Lawrence and Holyoke as such. In these two districts, which are dead last on the Boston Globe's MCAS rankings (Boston Globe, 2001a), 61% of 10th graders failed the math MCAS in 2001, and more than 65% of the students qualify for free or reduced-price lunch. By contrast, Brockton, with fewer math failures (38%) and fewer poor students (38%), while certainly deserving extra resources, does not have such extreme needs. Nevertheless, Brockton was on the Department's new list of high-need districts and Lawrence and Holyoke are not.

The second problem with the state's renewed commitment to high-need districts is that the new list of six such districts posted on the Department's web site in spring 2002 does not match another list of nineteen such districts that the Department provided at the same time to researchers it commissioned to evaluate the Bonus and MINT programs. (Note 4) Indeed, between 1998 and 2002, state officials referred to four different sets of high-need districts: 1) the thirteen districts named in the FY 1999 recruiting materials; 2) the 50 (unnamed) districts that then Associate Commissioner Alan Safran mentioned to a reporter in February 2001; 3) the list of 19 districts that DOE officials provided to Churchill et al. (2002); and 4) the six districts named in the FY 2002 recruiting materials.

The third problem with the state's new (spring 2002) commitment to high-need districts is that the state failed to fulfill this commitment. Fewer than half (40%) of the FY 2002 cohort of bonus recipients (N=43) began teaching in fall 2002 in the newly-designated set of six high-need districts. (Note 5) And 12% of this, the fourth cohort of Bonus recipients, are teaching in districts that scored in the top 50 on the 2001 MCAS.

Attrition

The bonus program has lost teachers at rates well above comparable national averages. I describe next the attrition rates for each cohort of bonus recipients and then discuss these findings as a whole.

First Cohort (FY 1999)

The first cohort of bonus recipients (N=59) began teaching fall 1999. In fall 2000, 20% (12 of 59) of this cohort did not return to the classroom, an attrition rate that is more than twice the national rate of 9% for first-year teachers (NCES, 1998). In the 13 high-need districts, the attrition rate was higher, with 31% (18 of 42) either stopping teaching or migrating to non-high-need areas (Fowler, 2001).

In fall 2001, seven more bonus recipients did not return to teaching, raising the attrition rate to 32% (19 of 59) after two years. Concurrently, 48% (20 of 42) of the first cohort who were initially placed in the high-need areas stopped teaching in these areas. This loss was even greater in particular districts. According to the Department's records, 91% (10 of 11) of the first cohort of bonus recipients in Chelsea stopped teaching, or left to teach elsewhere, after just two years.

After the third year, eight more members of this cohort stopped teaching, raising this group's attrition rate to 46% (27 of 59), more than twice the national 3-year attrition rate of 20% (NCES). And attrition for this cohort's high-need teachers rose to 55% (23 of 42), a loss of more teachers in three years than the nation's cities lose in five years (i.e., 50%) (Darling-Hammond & Schlan, 1996).

Second Cohort (FY 2000)

The second cohort of bonus recipients (N=105) began teaching in fall 2000. In fall 2001, 12% (13 of 105) of this group stopped teaching, with 18% (8 of 45) of those placed in the high-need areas either exiting the profession or migrating to non-high-need districts.

In fall 2002, 16 more members of this cohort left teaching, raising their 2-year attrition rate to 28% (29 of 105). Concurrently, five bonus recipients who had been teaching in high-need districts did not return to these districts, bringing the 2-year attrition rate for this group to 29% (13 of 45).

Third Cohort (FY 2001)

The third cohort of bonus recipients (N=105) began teaching in fall 2001. In fall 2002, 17% (18 of 105) of this cohort stopped teaching. Of the 37 bonus teachers working in high-need areas, 30% (11 of 37) did not return to teach in a high-need area.

Analysis

The preceding data, summarized in Table 6 below, indicates that each cohort of bonus recipients has sustained attrition rates that are above—usually well above—comparable national averages. While the United States loses 9% of new teachers after their first year of teaching, each cohort of bonus teachers exceeded this rate. Indeed, the first cohort's 1-year attrition rate (20%) more than doubled the national rate (9%), and the third cohort's first year losses (17%) came within one percentage point of doubling the national rate as well. Although the second cohort's 1-year attrition rate (12%) was closer to the national average, its 2-year rate (28%) exceeded the nation's 3-year rate (20%) by eight percentage points.

Table 6
New Teacher Attrition Rates

Teaching Population			**Percentage of new teachers who left after:**				
			1 year	**2 years**	**3 years**	**4 years**	**5 years**
Mass. Bonus Teachers	1st cohort (FY 1999)	All	20%	32%	46%	-	-
		High-need	31%	48%	55%	-	-
	2nd cohort (FY 2000)	All	12%	28%	-	-	-
		High-need	18%	28%			
	3rd cohort (FY 2001)	All	17%				
		High-need	29%		-	-	-
United States	All		*9%		**20%		
	Urban areas only						***50%

* (NCES, 1998); ** (Education Week, 2000); *** (Darling-Hammond & Schalan, 1996)

There is greater variation between the three cohorts' attrition rates for bonus teachers placed in high-need districts. In the first cohort, attrition was high, with 31% leaving after the first year, 48% after the second year, and 55% after the third. The second cohort's attrition rate for such teachers was not as high, though, with 18% leaving after one year, although this rate climbed to 29% after two years. However, the third cohort's 1-year attrition rate for high-need teachers, 30%, nearly equaled the first cohort's attrition rate for the same interval and exceeded the second cohort's 2-year attrition rate.

The attrition rates reported above are consistent with previous research which indicates that teachers who emerge from abbreviated preparation programs leave teaching sooner than those who emerge from longer preparation programs (Darling-Hammond, 2001). Indeed, given this fact, it may that the Bonus Program's attrition rate might have been even higher had all its participants been entirely new to teaching. Data obtained from the Department indicate, however, that a surprisingly high number of individuals who attended MINT had substantial prior educational experience. Specifically, 22% of MINT 2000 participants (more than 70% of them bonus recipients) indicated on a Department-sponsored survey that they had "attended a teacher training program (as part of their undergraduate or graduate education) before attending MINT." Another 5% indicated that they had substantial prior teaching experience, ranging from one year's teaching experience as a long-term substitute, to four years' experience teaching in both and public and private schools. Consequently, 27% of the FY 2000 MINT participants had substantial educational experience before attending MINT. It is possible, therefore, that the bonus recipients who were truly new to teaching left the classroom at even higher rates than reported above.

These high attrition rates entail, however, not just a loss of teachers, but a loss of training and bonus money. The state has now spent $921,250 ($748,000 for bonuses and $173,250 for MINT scholarships) for the 74 bonus recipients who have left the program. Ironically, Churchill et al. (2002) found that 76% of MINT graduates said they would have entered MINT without a bonus, a finding that concurs with other research (Liu, et al, 2000).

Program Effectiveness

One kind of evidence that Massachusetts officials offered in 2001 as proof of the Bonus Program's effectiveness was the large number of working teachers who have graduated from MINT. This number included not only Bonus recipients (62% of all MINT graduates), but also scholarship recipients (23%-30%) (Note 6), and other MINT graduates as well. For example, Commissioner Driscoll stated in June 2001 that "over the last three years, this program has put more than 450 teachers into classrooms. These are highly talented professionals who wouldn't be in schools today, if not for this initiative" (Massachusetts Department of Education, 2001c).

If accurate, this would be an impressive contribution over three years. However, analysis of Churchill et al.'s (2002) research indicates that the preceding statement inflates this contribution in a number of ways. First, the state produced not 450, but 444 MINT graduates between 1999 and 2001. Second, 44 of those graduates were no longer teaching in fall 2001. Third, the state was unable to confirm the teaching status of another 79 graduates. Consequently, 123 (28%) of MINT graduates had either stopped teaching or the state was unable to confirm whether they are teaching. Thus, the state may claim, at most, that 321 MINT graduates taught in the public schools in 2001-2002.

This would still be a substantial contribution if, as Commissioner Driscoll has claimed, these individuals would not otherwise have entered the teaching profession. But this is not the case. Churchill et al. (2002) found that, not all, but 61% of MINT graduates say they would not have entered the classroom without the fast-track approach. This further reduces the contribution of the MINT and Bonus programs to approximately 196 teachers who, to use Commissioner Driscoll's terms, "wouldn't be in schools today, if not for this initiative"(Massachusetts Department of Education, 2001c).

The effectiveness of the Bonus Program should be judged, however, not by the number of such teachers "in schools today," but by the number of such teachers in high-need schools or school districts. Why? Because one of "the three major goals for the MINT/Signing Bonus program [is to] .... address teacher shortages in high-need schools" (Churchill et al., 2002, p. 6). The Bonus Program falls short in this area, though, because Churchill et al. (2002) found that, during the 2001-2002 academic year, just 34% of MINT graduates were teaching in the 19 high-need districts that the Department identified for these researchers, which means that approximately 67 of the 196 teachers mentioned above taught in such districts. Consequently, the Bonus and MINT programs together have brought to state-designated high-need schools a small number of individuals who would not otherwise have gone into teaching.

Another kind of evidence that policy makers have offered regarding program effectiveness is principals' responses to survey questions about MINT graduates. For example, when Commissioner Driscoll issued a statement defending the Bonus and MINT programs in February 2001, he wrote that "90% of their principals want to hire more" Bonus recipients and MINT graduates (Massachusetts Department of Education, 2001b). The preceding statement does not capture accurately, however, what principals said. In the first MINT evaluation, the source of this statistic, the author wrote that "most principals noted that hires are made following an interview process" (Massachusetts Department of Education, 2000b, p. 9). This comment implies that principals were open to hiring, but not necessarily actively seeking out, more MINT graduates. Findings by Churchill et al. (2002) confirm this implication. When they asked principals how they would consider hiring more MINT graduates, 10.6% said they would consider them "with preference;" 10.7% said "with reservations [9.4%] or not at all [1.3%];" and 79% said "the same as everyone else." In other words, principals who have supervised Bonus and MINT graduates are neither eager nor reluctant to hire more MINT graduates, rather they are willing to consider them as they would other traditionally-trained teachers.

Finally, state officials have pointed to principals' comparatively favorable ratings of MINT-trained versus traditionally-trained teachers. For example, they reported that "71% of principals who were interviewed ... rate their MINT graduate/Bonus Recipient(s) as average (42%), above average (19%), or well above (10%) average compared to all of the teachers at their school" (Massachusetts Department of Education, 2000c).

The source of the preceding statistic, the 2000 MINT and Bonus evaluation, lacks credibility, in part because it failed to report either the number or percentage of principals' unfavorable ratings for survey questions of this kind. In order to ascertain this information, I requested all of the original survey data from the Department and calculated the percentages for the statistic cited above. I found that the evaluator had calculated this statistic incorrectly: 64% (not 71%) of principals had rated bonus teachers average or above. Further, when I weighted the principals' ratings by the number of teachers they supervised, I found that 60% of Bonus teachers were rated as average or above. When these percentages are aggregated in the other direction, though, the same percentage of teachers, 60%, were rated as average or below.

Further doubt surrounds this evaluation because it wrongly implied that the attrition rates of the first cohort of first-year teachers (FY 1999) was less than 15%, when it was actually 20%. Moreover, its author was, at the time she wrote the evaluation, a contract employee of the Department who went on national recruiting trips for the Bonus Program and earned approximately $150,000, over three years, while working for this and other related Department programs. This is not to say that the evaluator did not strive to be objective, but it is to say that few would consider a contract employee whose future employment depends on a) the continued existence of the programs which she is simultaneously working for and evaluating, and b) the good will of her immediate supervisors whose performance she is indirectly judging, to be independent. Indeed, it is doubtful that the Department, which calls for school districts to produce "outside evaluation reports" (Massachusetts Department of Education, 2000d, p. 2) as evidence of effective performance, would consider an evaluation produced by such an employee to be an "outside" evaluation.

Due to the factors cited above, the 2000 Bonus Program evaluation is not credible. Fortunately, Churchill et al. (2002) have conducted an evaluation that is far more credible. The authors of this Department-commissioned report a) are independent and b) surveyed far more individuals (including MINT graduates, supervising principals and trainers) than the author of the 2000 evaluation. Although it is not practical to summarize the many findings of this wide-ranging report here, it is pertinent to note that the authors also asked principals to compare MINT-trained and traditionally-trained teachers. They found that "overall, the principals were very positive about the quality of the MINT graduates" (Churchill et al., 2002, p. 15) when comparing MINT graduates' with other traditionally-teachers in the following areas: a) content knowledge, b) ability to employ effective instructional strategies, c) ability to work with students with special needs, and d) classroom management skills.

It is not clear, though, whether, and to what extent (if any), these ratings varied according to where graduates' taught (high-need vs. suburban districts). Absent such details, it is not possible to say whether principals in high-need districts were more or less satisfied with the performance of the MINT graduates they supervised, particularly those with little or no prior educational experience.

It is pertinent to note, though, that this Department-commissioned evaluation questioned whether the state's seven-week fast-track approach to teacher preparation was sufficient for high-need, largely urban districts. The authors recommend, among other items, that that the state "consider whether MINT needs to be changed fundamentally if it is to meet its current goal of serving high-need districts. The demands of urban teaching are such that a Bonus and a seven-week summer session simply may not be sufficient to adequately prepare significant numbers of high-quality teachers who will stay" (Churchill et al., 2002, p. 39). They go on to suggest that a variety of one-year apprenticeship type programs may be more suitable.

The ultimate comment on the Bonus Program's effectiveness, though, is not what the state's policy makers have said about it but what they have done to it. Their recent radical revision of this initiative is an implicit admission that the program suffered from chronic problems. Although state officials have not specifically named these problems, the current research has identified a number of issues that have beset this initiative since its inception. As I discuss next, these problems raise doubts about federal policy makers' claims that fast-track certification programs will help the country produce more and better teachers in the coming years.

Implications for State and National Policy

The US Department of Education is a strong proponent of fast-track certification initiatives. It claims that if more such programs are adopted nationally, this will open up a new, previously untapped pool of teachers to the country and thereby deliver the following five benefits to the nation's schools: 1) more teachers, 2) more urban teachers, 3) more minority teachers, 4) higher teacher retention rates, and 5) better teachers (U.S. Department of Education, 2002). Since Massachusetts's Bonus and MINT programs were a) designed in accordance with the principles and practices the federal government is exhorting other states to adopt, and b) managed for the most part by an organization (the new Teacher Project) that the Bush administration has lauded for its ability to implement such initiatives (The White House, 2001a, 2001b, 2002), it is appropriate to consider to what extent Massachusetts has realized each of these benefits.

1. More Teachers

   Approximately 196 individuals, MINT graduates who would not have otherwise gone into teaching, taught in Massachusetts's public schools in 2001-2002. This provided a moderately positive boost to the state's supply of teachers, comparable to the number of working teachers that one of the state's mid-sized schools of education would produce over a similar three-year period.

2. More Urban Teachers

   The first three cohorts of the Bonus Program produced approximately 67 working teachers who would not otherwise have entered teaching and who taught in the state's high-need school districts in 2001-2002. This represents a negligible increase in the state's supply of urban teachers.

3. More Minority Teachers

   Department records (obtained via a FOIA request) indicate that 9% of the 2002 MINT trainees are minority. This is one percentage point lower than the 10% of new teachers hired in Massachusetts in 1999 who were from minority groups (Massachusetts Department of Education, 2000e). Churchill et al. wrote of the Bonus Program's minority teacher recruitment, "there has been negligible success recruiting minority candidates, to date" (2002, p. 37).

4. Higher Teacher Retention Rates

   According to the US Department of Education, "nationwide, about 85 percent of teachers certified through alternate routes remain in the classroom five years later, demonstrating that truncated training programs with highly qualified candidates do not result in those same teachers leaving the profession early in their careers" (2001, p. 16).

   This has not occurred in Massachusetts, however, where only 54% of the first cohort of bounus recipients remain after three years, 72% of the second cohort after two years, and 71% of the third cohort after one year. These retention rates were even lower in state-designated high-need districts, with 45% of the first cohort remaining after three years, 72% of the second after two years, and 71% of the third after one year.

5. Better Teachers

   To date, the Bonus Program has not offered any objective, test-based evidence (of the sort that the federal government and most state governments now require schools to produce) to indicate that its teachers are raising children's academic achievement. Massachusetts has produced, however, survey data which indicates that principals rated MINT graduates' performance favorably, when compared to traditionally-trained teachers. It is not clear, though, whether principals' ratings varied either by a) the extent of the teacher's prior educational experience or b) the nature of the teacher's placement (urban vs. suburban). Nor is it clear whether these ratings predict improved student performance.

Conclusion

The Massachusetts Signing Bonus Program for New Teachers has failed to produce in Massachusetts the positive outcomes that federal policy makers promise such programs will produce in the nation, even though this program was modeled after Teach for America and critical parts of it were designed and often managed by the New Teacher Project—two related organizations that the Bush administration has repeatedly and effusively praised for their ability to both design and run programs of this type. This does not necessarily mean, of course, that fast-track alternative certification will not work, either in Massachusetts or elsewhere. It strongly implies, however, that simply embracing the two core principles the federal government has endorsed for such programs—1) recruit high-achieving candidates and 2) get them quickly into the classroom (US Department of Education, 2002)—will fail to deliver the benefits the Bush administration is promising.

Officials should respond to the current research by resisting calls to embrace rapid certification. It would be unwise to devote extensive funding to an approach that has produced, in Massachusetts, such low numbers of urban teachers and such high numbers of exiting teachers, all at a cost of more than four million dollars. Given the current teacher shortages in many states, though, it is understandable that policy makers at all levels of government are increasingly willing to experiment with alternative forms of teacher preparation. It is critical, however, that when officials fund such initiatives, they do so with a spirit of experimentation—and with the rigor, transparency, and objectivity that characterize proper experimentation.

Policy makers must insist on ongoing and independent evaluations of any experimental program, evaluations that track participants' characteristics, such as prior educational experience, eventual teaching placements, retention rates, and more. Absent rigorous collection and independent analysis of such data, the public will not know whether such initiatives are effective. Independent analysis of carefully collected data is critical if other states are to avoid what happened in Massachusetts, where state officials have a) offered inaccurate information regarding the number and characteristics of MINT graduates working in schools, and b) issued an audit in 2000 that inflated the program's favorable statistics (principals' ratings) and underreported its unfavorable statistics (the first-year attrition rate). This inaccurate information helped convince some that Massachusetts is developing a model of teacher preparation for other states to follow. When independent data are brought to bear on this program, however, this initiative looks less like a model of teacher preparation worth copying and more like an expensive quick-fix that has failed to solve a complex problem.

References

Archer, J. (2002, December 4) Education Week. Mass. Bonus program to favor ed schools. Available from: http://www.edweek.org/ew/ewstory.cfm?slug=14bonus.h22

Associated Press (2001, February 16). Teacher bonus program criticized. *Cape Cod Times*.

Boston Globe (2001a). Comparing the districts. Available from: http://www.boston.com/mcas/scores2001/globe_districts_rank01.htm

Boston Globe (2001b). Top-ranked 10th grade districts. Available from: http://www.boston.com/mcas/scores2001/10th_district_rank01.htm.

Churchill, A., Berger, J., Brooks, C,. Effrat, A., Grifin, L., Magouirk-Colbert, M., McDermott, K., Sharick, R., & Sheehan, A. (2002, May 21). *Revised interim report: The Massachusetts Institute for New Teachers and Master Teacher/NBPTS Programs*. Amherst, MA: Center for Education Policy—University of Massachusetts.

Commonwealth of Massachusetts (2001, July 26). Swift urges alternative teaching paths: Praises MINT teachers for commitment. Available from: http://www.state.ma.us/gov/press/pr072601shrewsburyhigh.htm.

Darling-Hammond, L. (2002, September 6). Research and rhetoric on teacher certification: A response to "Teacher Certification Reconsidered," *Education Policy Analysis Archives, 10*(36). Retrieved April 8, 2003 from http://epaa.asu.edu/epaa/v10n36.html

Darling-Hammond, L. & Schlan, E. (1996). Who teaches and why? Dilemmas of building a better profession for twenty-first century schools. In J. Sikula (Ed.), *Handbook of research on teacher education*. (2nd. ed., pp. 67-101). New York: MacMillan.

Education Week (2000). *Quality Counts 2000: Who Should Teach?*

Education Week (2002). *Quality Counts 2002: Who Should Teach?*

Ferdinand, P. (1999, February 2). $20,000 bonus prompts a rush for Massachusetts teaching jobs. *The Washington Post*, p. A02.

Fowler, R. C. (2001, February) *An analysis of the recruitment, preparation, attrition, and placement of the Massachusetts signing bonus teachers.*

Available from: http://www.massteacher.org/html/Public_area/Of_Interest.html.

Hayward, E. (2000, July 6). Bay State teacher recruiting cheered as model for the nation. *Boston Herald*, p. P 26.

Hayward, E. (2002, January 30). Applications for teacher signing bonuses drop. *Boston Herald*, p. P23.

Liu, E., Kardos, S., Kauffman, D. Peske, H., & Johnson, S. (2000, April). "Breaking even:" Incentives, rewards, and the high costs of choosing to teach. Paper presented at the annual meeting of the American Education Research Association, New Orleans, LA.

Magee, M. (1999, January 16) New teachers become hot commodity in nationwide hunt. *San Diego Union-Tribune*, p. B-1.

Massachusetts Department of Education (1998, September 15). 12 to 62 Teacher enhancement legislation: timeline for proposed roll out. Available from: www.doe.mass.edu/eq/news98/1262tmln.html.

Massachusetts Department of Education (1999a) Board of Education 1999 Annual Report. Available from: http://www.doe.mass.edu/boe/annual99/toc.html.

Massachusetts Department of Education (1999b) Massachusetts Signing Bonus Program: questions and answers.

Massachusetts Department of Education, (1999c, May 24) "20,000 Signing Bonus Press Conference, May 24, 1999" (videotape).

Massachusetts Department of Education (2000a, March 21) Commissioner's Update. Available from: http://www.doe.mass.edu/mailings/2000/321/cm032100.html

Massachusetts Department of Education (2000b). Massachusetts Institute for New Teachers Report. Available from: http://www.doe.mass.edu/eq/mint/history.html.

Massachusetts Department of Education (2000c, December 26). Massachusetts School Principals Give High Ratings to Signing Bonus Teachers. Available from: http://www.doe.mass.edu/news/news.asp?id=167

Massachusetts Department of Education (2000d). District performance evaluation instrument. Available from: http//www.doe.massedu/ata/dpe00/instrument/section_4.html.

Massachusetts Department of Education (2000e). Education Reform Staff System Report: August 1, 1999-Ocober1, 1999.

Massachusetts Department of Education (2001a, February 15). Driscoll Driscoll, Massachusetts Commissioner of Education: A statement regarding the teacher signing bonus program and MINT (Massachusetts Institute for New Teachers) programs. Available from: http://www.doe.mass.edu/news/archive01/su021601_2.html

Massachusetts Department of Education (2001b, June 1) 230 to get accelerated training to teach this fall. Available from: http://www.doe.mass.edu/news/news.asp?id=295.

Massachusetts Department of Education (2002a, November 20). Colleges and universities tapped to nominate future bonus recipients under new guidelines. Available from: http://www.doe.mass.edu/news/news.asp?id=1292

Massachusetts Department of Education (2002b). The Massachusetts Signing Bonus Program. Available from: http://eq.doe.mass.edu/mint/overview.html.

National Council on Teacher Quality (Ed.) (2002) Unconventional Recruiting: Places to Watch. Available from: http://www.nctq.org/issues/recruiting.html

New Teacher Project (Ed.). (2002) Project highlights. Available from: http://www.tntp.org/cli/hlt.html

NCES (1998). *The condition of education 1998*. Washington, DC: U.S. Department of Education.

Salzer, J. (2/4/02). Budget crunch may hurt teacher incentive plan. *The Atlanta Journal-Constitution*.

School Board News (2001, February 20). Teacher bonuses paying off in Massachusetts. Available from: http://www.nsba.org/sbn/01--feb/022001-5.htm.

Setera, K. (Segment Producer). (2001, February 15). Teacher Bonuses. *Fox 25 News at Ten: WXFT*. A transcript of this broadcast is available at http://www.fox25.com/. Click on "Fox25 Undercover" and then "Teacher Bonuses."

Silberman, T. (2001, July 3). "Code Red" in teaching. *Charlotte News and Observer*.

Turner, J. (February 1, 2001). Teacher recruitment is top priority in Western Virginia. *The Roanoke Times*, p. B4.

U. S. Department of Education (2002). *Meeting the highly qualified teachers challenge: The Secretary's annual report on teacher quality*. Washington, DC. Available: from: http://www.edgov/offices/OPE/News/teacherprep/index.html.

Vaishnav, A. (2001, February 16). Cities said to lose teacher recruits, *The Boston Globe*, p. B3.

Vaishnav, A. (2002, March 21).Teachers' signing bonuses cut back. *The Boston Globe*, p. B1.

The White House. (2001a, April 3) Remarks of Laura Bush at Senate Spouses Luncheon. Available from: http://whitehouse.gov/firstlady/news-speeches/speches/fl20010403-2.html.

The White House. (2001b, April 3) Remarks of Laura Bush to DC Teaching Fellows Launch. Available from: http://whitehouse.gov/firstlady/news-speeches/speches/fl20010222.html.

The White House. (2002, February, 26). Remarks by Mrs. Bush—Transition to Teaching/New Teacher Project. Available from: http://whitehouse.gov/firstlady/news-speeches/speches/print/fl20020226.html

Zernike, K. (2000, August 24). Less Training, more teachers: new math for staffing classes. *New York Times*, p. A1.

About the Author

Clarke Fowler
Education Department
Salem State College
Salem, MA 01970
(978) 542-7041
clarke.fowler@salemstate.edu

Dr. Clarke Fowler, an Associate Professor in the Education Department at Salem State College, has published written articles on socio-moral and cognitive development. His most recent articles have focused on the Massachusetts Teacher Tests, Title 2 of the Higher Education Reauthorization Act of 1998, and Massachusetts's fast-track teacher certification initiatives.

Notes

1. To calculate these costs, I obtained, via a series of Freedom of Information requests, all recruitment-related travel documents from the Department. To calculate personnel costs, which the Department did not include in its recruiting budget, I identified the employees who went on each trip and then used salary schedules to estimate how much these individuals earned while recruiting.
2. The Department's initial set of recruiting materials explicitly stated, in a Question-and-Answer format, that bonus recipients would teach in thirteen high-need urban districts: "Q: What are the possible placement sites? A: Boston, Brockton, Cambridge, Chelsea, Fall River, Framingham, Holyoke, Lawrence, Lowell, Lynn, New Bedford, Springfield, and Worcester" (Massachusetts Department of Education, 1999b).
3. Note 2 (above) contains the wording of the 1999 recruiting materials. Commissioner Driscoll reaffirmed this commitment orally during a May 1999 press conference, when he announced the awarding of the first round of bonuses: "We're working with the urban areas in Massachusetts where these recipients will be placed." Later, he said, "We have commitments from all of the major urban areas— Boston, Cambridge, Chelsea, Fall River, New Bedford, Holyoke, Springfield, Lawrence, Lowell—so we have commitments from all of the schools districts. It's just a matter of matching them up" (Massachusetts Department of Education, 1999c). Finally, the Massachusetts Board of Education' s1999 Annual Report says: "New teachers begin their four-year commitment in the summer in an intensive, seven-week teacher-training institute, and groups of four to five began teaching in thirteen high needs Massachusetts public school districts in the fall of 1999" (Massachusetts Department of Education, 1999a, p. 20).

4. Churchill et al. (2002) reported that "The Department of Education has identified nineteen districts as being high-need, based on overall number of students, percent of students qualifying for free- or reduced-price lunch, and MCAS scores" (p. 26). These districts are Boston, Brockton, Cambridge, Chelsea, Chicopee, Fall River, Haverhill, Holyoke, Lawrence, Lowell, Lynn, New Bedford, Pittsfield, Revere, Salem, Somerville, Springfield, Taunton, and Worcester.
5. Whereas the state sought to recruit 125 bonus teachers in 2000 and 2001, in 2002 they planned on recruiting no more than 50. The Boston Herald reported that, due to the poor performance of the bonus program's endowment, much less money was currently available to pay for bonuses (Hayward, 2002). At 43, the FY 2002 cohort of bonus teachers is the smallest cohort the Bonus program has placed in schools to date.
6. I report a range of scholarship recipients because the Department's records are incomplete: officials are unsure of the award status of 39 of their MINT graduates. That is, the Department did no know whether the state, the individual, or a school district paid the $2,250 tuition for these individuals to attend MINT.

Copyright 2003 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

This article has been retrieved 1820 times since April 29, 2003

Education Policy Analysis Archives
Volume 11 Number 14 **April 29, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Educational Aspirations and Postsecondary Access and Choice: Students in Urban, Suburban, and Rural Schools Compared

Shouping Hu
Seton Hall University

Citation: Hu, S. (April 29, 2003). Educational aspirations and postsecondary access and choice: Students in urban, suburban, and rural schools compared. *Education Policy Analysis Archives, 11*(14). Retrieved [date] from http://epaa.asu.edu/epaa/v11n14/.

Abstract

Using data from the National Education Longitudinal Study of 1988 (*NELS: 88*), this study examines educational aspirations and postsecondary access and choice by students in urban, suburban, and rural schools. In addition, this study raises issues with the methods in postsecondary educational research by using students in different grades (8th, 10th, and 12th grades) as baseline populations to compare educational outcomes. The results indicated that students in urban schools were comparatively disadvantaged in the early years in schooling in terms of postsecondary access but appeared to be enrolled in postsecondary institutions at similar percentages as their suburban counterparts, if they made it to later years in K-12 schooling. For those students in urban schools who went to college, higher percentages were enrolled in private institutions and four-year colleges. Students in rural schools were consistently disadvantaged in postsecondary aspirations and enrollment, compared to students in other schools.

Introduction

Educational researchers and policy analysts have been interested in educational quality received by students in different types of schools for years. For instance, Anyon (1997) revealed how urban schools were failing students in her study on the Newark public school system in New Jersey. DeYoung (1987) reviewed research on American rural education and suggested the challenges rural schools were facing in educating school children. Researchers also suggested that students in different types of schools have different levels of academic achievement and educational attainment due to the disparity in family and school resources (McDonough, 1997; Roscigno & Crowley, 2001) and student learning opportunities (Adelman, 1999). However, very few studies examined how postsecondary opportunity was distributed among students in different types of schools as classified as urban, suburban, and rural schools, even though postsecondary readiness and participation are among the most important issues in the state and federal higher education policy arena (Heller, 2001; National Center for Public Policy and Higher Education, 2001). This is a troubling situation in light of the increasing calls for strengthening the connection between K-12 schools and higher education institutions (Maeroff, Callan, & Usdan, 2001; Stampen & Hansen, 1999). As researchers (Hannaway & Talbert, 1993; Shouse, 1998) have found that school context is important in identifying the effective practices in schools, school context would logically also be important in the efforts to bridge K-12 schools and postsecondary institutions (Maeroff, Callan, & Usdan, 2001).

Furthermore, when studying postsecondary access and choice, researchers need to carefully consider how to conceptualize the ideal of equal educational opportunity (Burbules, 1999; Howe, 1997), because there is large disparity in the dropout rates among students in urban, suburban, and rural schools (Roscigno & Crowley, 2001; Rumberger & Thomas, 2000). For example, using NELS data to examine enrollment decisions by 12th graders, Perna (2000) found that students in urban and rural schools actually were more likely to go to college than their suburban counterparts, controlling for student characteristics and a series of other factors. Student transition from the 12th grade to college is a research area of significant policy importance. However, it alone may not adequately address the disparity in postsecondary opportunity because some disadvantaged students dropped out of school before reaching the 12th grades (Alexander, Entwisle, & Kabbani, 2001; Orfield, 1988).

Purpose

This study focuses on the critical transition points in student pathways to postsecondary education. Student pathways to college are considered as a multi-stage process including educational aspiration formation, academic preparation, and actual enrollment in college, a process could start as early as the 7th grades (Choy, Horn, Nuñez, & Chen, 2000; Terenzini, Cabrera, & Bernal, 2001). Based upon previous research (e.g., Cabrera & La Nasa, 2001; Hossler & Gallagher, 1987; St. John, Asker, & Hu, 2001; Terenzini, Cabrera, & Bernal, 2001) and different conceptions of the ideal of equal educational opportunity (Burbules, 1999; Howe, 1997), this study examines student educational aspirations and access to postsecondary education by using 8th, 10th, and 12th graders as baseline populations. Further, for those enrolled in postsecondary education after two years out of higher school, this study contrasts their college destinations (four-year *vs.* two-year, public *vs.* private) with respect to their 12th grade school origin (urban, suburban, and rural schools).

Specifically, this study intends to answer the following questions using nationally representative samples:

1. How do educational aspirations measured in the 10th grade differ for students in urban, suburban, and rural schools with the 8th and 10th graders as the baseline population?
2. How does postsecondary access measured two year after high school differ for students in urban, suburban, and rural schools with the 8th, 10th, and 12th graders as the baseline population?
3. How does postsecondary choice for enrolled students differ with respect to their origins in urban, suburban, and rural 12th grade schools?

Method

Data

Data used for this study were from the National Educational Longitudinal Study 1988 (NELS: 88). NELS was sponsored by the US Department of Education National Center for Educational Statistics (NCES) to survey a cohort of students in the 8th grade (base year in1988), the 10th grade (first follow-up in 1990), the 12th grade (second follow-up in 1992), and two years after high school graduation (third follow-up in 1994). Recently, the NCES released the fourth follow-up survey of 2000. In order to make the sample representative for different baseline populations, different weights were used in this study (Huang, Salvucci, Peng, & Owings, 1996).

Variables and Analysis

To examine whether there are differences in student educational aspirations and postsecondary access and choice, three outcome variables are used in this study on the basis of their socioeconomic significance in individual mobility (Pascarella & Terenzini, 1991; Terenzini, Cabrera, & Bernal, 2001).

The first outcome variable is student educational aspirations measured when students were in the 10th grade. F1 (the first follow-up, similar connotations for F2 and F3) panel weight (F1PNLWT) and F1 questionnaire weight (F1QWT) were used to project baseline population of the 8th and 10th graders respectively. Since "educational aspiration" as a construct has been tested as important in understanding individual college access and choice by the literature in sociology and education, and the survey items in NELS were accepted as valid measures on this construct, it was selected as one outcome variable in this study. In fact, Hearn (1992) argued that using educational aspiration as a construct has been "near-paradigmatic" (p. 662) in postsecondary enrollment research, although he acknowledged some potential issues with this construct.

The second one is student postsecondary access measured two years after high school graduation. The final outcome variable is postsecondary institutional types chosen by enrolled students measured two years after high school graduation. F3 panel weight (F3PNLWT), F3F1 panel weight (F3F1PNWT), and F3F2 panel weight (F3F2PNWT) were used to project baseline population of the 8th, 10th, and 12th graders respectively for the latter two outcome variables. "Access" deals with whether students go to college or not and "choice" deals with where students go to college. Both have been considered as important outcome variables in postsecondary policy studies (McPherson & Schapiro, 1991).

The independent variable in this study is school location (school urbanicity as in the NELS data set) as classified as urban, suburban, and rural to reflect the sample school's metropolitan status. Urban represents central city, suburban represents areas surrounding a central city within a county constituting the Metropolitan Statistical Area (MSA), and rural represents areas outside MSA. The composition of students in the samples during the 8th, 10th, and 12th grades was essentially as follows: slightly lower than 1/3 in urban schools, slightly higher than 1/3 in suburban schools, and about 1/3 in rural schools.

This study was a descriptive analysis of a national database to indicate the unequal postsecondary opportunity by students in urban, suburban, and rural schools in their postsecondary educational aspirations, access, and choice. Cross tabulations were used to illustrate the overall differences in outcome variables by school type with respect to different baseline populations.

Results

Aspirations

Because there were no substantial differences in educational aspirations by using the 8th and 10th graders as baseline populations, Figure 1 only presents student educational aspirations with respect to school location for 10th graders as baseline population. Differences in educational aspirations by students in urban, suburban, and rural schools were evident. Higher percentages of students in rural schools had aspirations for high school or below (16.6% for rural in contrast to 11.0% for urban and 10.6% for suburban schools) and two year college education (33.1% for rural in contrast to 27.1% for urban and 29.3% for suburban schools), and lower percentages of rural students had aspirations for four year college education (28.2% for rural in contrast to 30.8% for urban and 32.9% for suburban schools) and graduate education (22.0% for rural in contrast to 31.1% for urban and 27.3% for suburban schools). There were no substantial differences in educational aspirations for students in urban and suburban schools, although it appeared that slightly higher percentage of urban students had aspirations for graduate education.

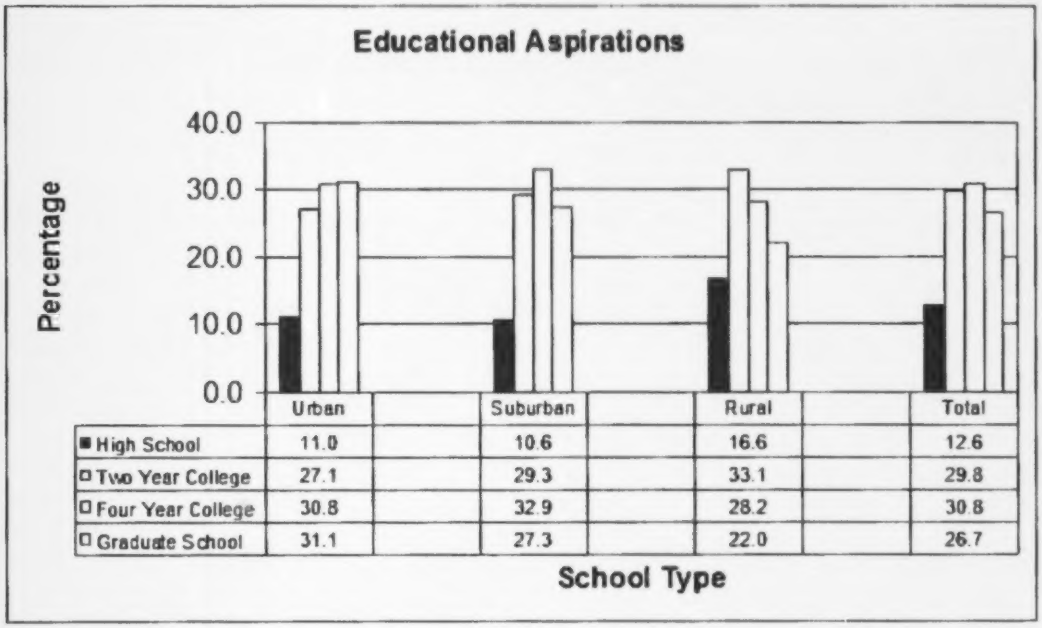


FIGURE 1. Educational Aspirations in the 10th Grade by Students in Urban, Suburban, and Rural Schools Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 - 1992.

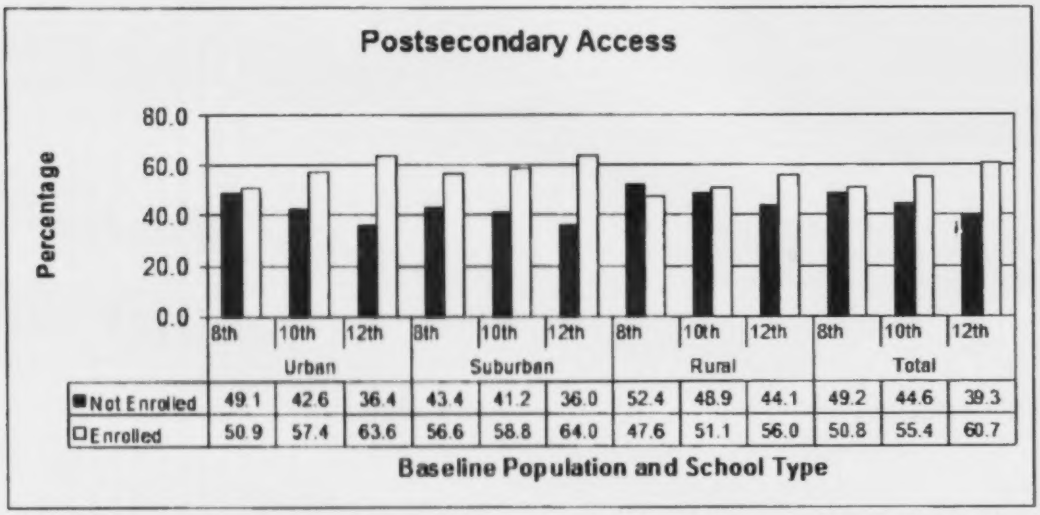
Access

As presented in Figure 2, student access to college was analyzed by comparing students who went to college to those who did not go to college by October 1992, two years after high school graduation. Student enrollment status in college by school location was analyzed using the 8th, 10th, and 12th graders as the baseline populations.

First, the percentage of student enrollment in postsecondary education increased, when the 8th, 10th, and 12th graders were used as baseline populations respectively. For students in urban schools, the enrollment rates increased from 50.9% to 57.4% and 63.6%. For students in suburban schools, the enrollment rates increased from 56.6% to 58.8% and 64.0%. For students in rural schools, the enrollment rates increased from 47.62% to 51.1% and 56.0%. This is understandable because some students may drop out during the middle school and high school schooling process.

Second, smaller percentages of students in rural schools were enrolled in postsecondary institutions, no matter which baseline population was used. When the 8th graders were used as the baseline population, the enrollment percentage for students in rural schools was 47.6%, in contrast to 50.9% in urban schools and 56.6% in suburban schools. When the 10th graders were used as the baseline population, the enrollment percentage for students in rural schools was 51.1%, in contrast to 57.4% in urban schools and 58.8% in suburban schools. When the 12th graders were used as the baseline population, the enrollment percentage for students in rural schools was 56.0%, in contrast to 63.6% in urban schools and 64.0% in suburban schools.

Third, although smaller percentages of students in urban schools were enrolled in college than their suburban counterparts when the 8th graders were the baseline population (50.9% *vs.* 56.6%), there were virtually no differences in the percentages of postsecondary enrollment by students in urban and suburban schools when the 12th graders were used as baseline population (63.6% *vs.* 64.0%).



	Urban 8th	Urban 10th	Urban 12th	Suburban 8th	Suburban 10th	Suburban 12th	Rural 8th	Rural 10th	Rural 12th	Total 8th	Total 10th	Total 12th
■ Not Enrolled	49.1	42.6	36.4	43.4	41.2	36.0	52.4	48.9	44.1	49.2	44.6	39.3
☐ Enrolled	50.9	57.4	63.6	56.6	58.8	64.0	47.6	51.1	56.0	50.8	55.4	60.7

FIGURE 2. Postsecondary Access by October 1992 by Students in Urban, Suburban, and Rural Schools Using the 8th, 10th, and 12th Graders as Baseline Populations
Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 Data Analysis System.

Choice

Figures 3 and 4 present results about postsecondary destinations for students who made it to the stage of postsecondary education. The population, therefore, was students who were enrolled in postsecondary institutions two years after high school graduation. The questions here were, for students who successfully reach the level of postsecondary education in different type of 12th grade schools (urban, suburban, and rural), what was their distribution in different types of postsecondary institutions?

Two major findings are worth reporting. First, for those who managed to go to college, relatively larger percentages of students in rural schools were enrolled in public institutions (78.5%), while relatively smaller percentages of students in urban schools were enrolled in public institutions (67.9%), and the percentages for students in suburban schools were in between (75.4%) (Figure 3). Second, relatively larger proportions of students in urban schools were enrolled in four-year college (60.8%), while there was no substantial difference between student in suburban (56.9%) and rural schools (56.4%) (Figure 4). Further analysis (not tabled) suggests relatively larger percentages of students in urban schools (24.4%) were enrolled in private not for profit four-year colleges than students in suburban (18.5%) and rural schools (16.3%).

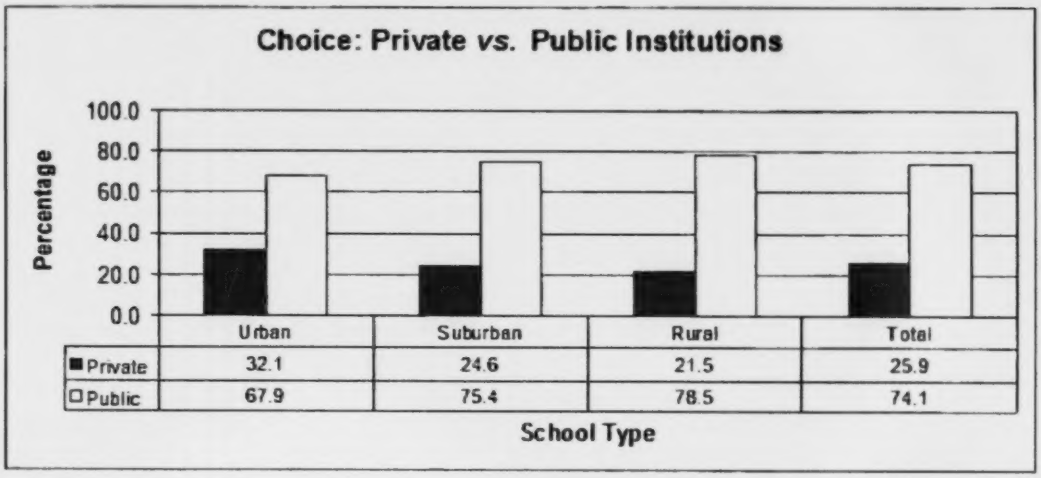


FIGURE 3. Choice of Private *vs.* Public Institutions by Students in Urban, Suburban, and Rural Schools Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 Data Analysis System.

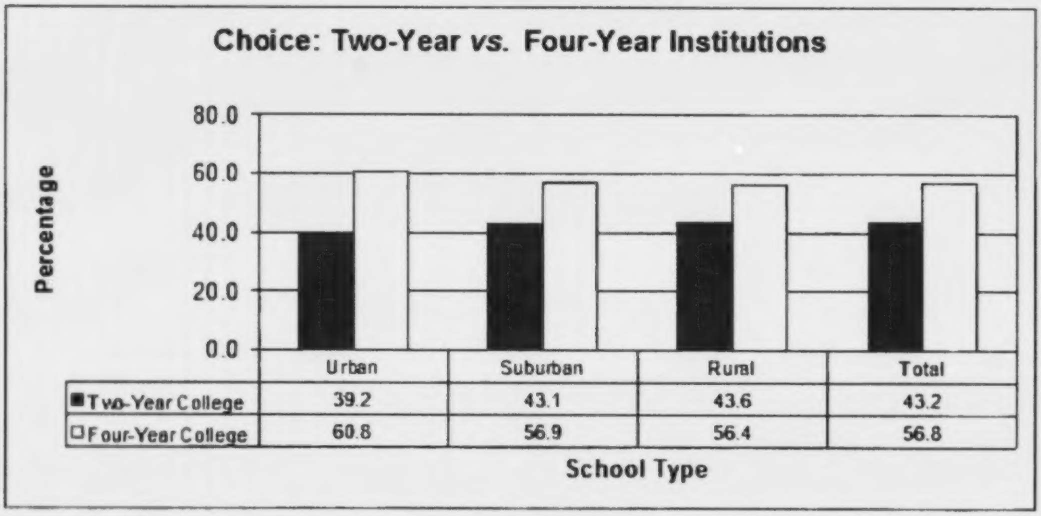


FIGURE 4. Choice of Two-Year *vs.* Four-Year Institutions by Students in Urban, Suburban, and Rural Schools Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 Data Analysis System.

Discussion

The influential report *Measuring Up 2000* issued by National Center for Public Policy and Higher Education (2001) identified preparation for and participation in postsecondary education among the most important postsecondary policy issues. This study examines these important policy issues in analyzing student pathways to postsecondary education—student educational aspirations, access, and choice in postsecondary education. Keeping in mind that this study is descriptive in nature, it has important policy implications in the following aspects.

First, this study examines the condition of school-location related unequal postsecondary opportunity. Compared to what we already know about postsecondary opportunity by students of different background such as gender, race/ethnicity, and socioeconomic background, we have little understanding and even some misunderstanding about how postsecondary opportunities were distributed among students in urban, suburban, and rural schools. This study offered an account on the condition of postsecondary opportunity among students in these three types of schools, using nationally representative samples. The results from this study suggest a potential new dimension of unequal educational opportunity—the location of the school. Specially, the consistent patterns of lower-level of educational aspiration, access, and choice by students in rural schools call for policy attentions. Policy makers need to consider policy interventions targeted toward to schools in different locations to promote postsecondary educational opportunity.

Secondly, this study raises questions about the conception of equal educational opportunity and related analytical methods. The results revealed that the unequal educational opportunity along the line of school location operates differently in different stages of student educational career. Although the 8th graders in urban schools are at smaller percentages of going to college than their suburban counterparts, 12th graders in urban schools, however, are at virtually equal percentages of being enrolled in postsecondary education, and even at higher percentages of going to private four-year colleges. This suggests that early interventions that can help student make to the later stage of K-12 schooling could be particularly effective strategies in promoting postsecondary educational opportunities for students in urban schools. Clearly it is important to examine the transition from the 12th grade to postsecondary education, but it alone may not be adequate to address policy concerns on equal postsecondary opportunity. It is important to track student progress through their educational career to promote equal educational opportunity.

Finally, this study provides important insights for future research. First, the combination of different conceptual and analytical frameworks will help researchers gain a full understanding about student postsecondary opportunity. Secondly, multivariate analyses that take into considerations of variables concerning student and school characteristics will help unravel the underlying process and factors related to the unequal educational opportunity for students in different types of schools. For instance, the school-location related inequality in educational opportunities might well be the consequences of the level of family poverty in different locales and the unequal offerings of learning opportunities (e.g., AP courses in high school) in different schools (Adelman, 1999; Hebel, 1999). Geography may also operate as a mediating mechanism by influencing the structure, decisions, and socialization opportunities in different communities and schools, which will then shape individual opportunities and educational choices (Coleman, 1988; Gamoran, 1987; McDonough, 1997; Roscigno & Crowley, 2001; Smith, Beaulieu, & Seraphine, 1995). Further exploration in these directions would be able to provide insights for more effective and implementable K-16 connection strategies.

Note

An earlier version of this paper was presented at the Annual Meeting of the American Educational Research Association (AERA), New Orleans, 2002. The author wants to thank anonymous reviewers of *EPAA* for their helpful comments.

References

Adelman, C. (1999). *Answers in the toolbox: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: US Department of Education, Office of Educational Research and Improvement.

Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *Teachers College Record, 103*, 760-822.

Anyon, J. (1997). *Ghetto schooling: A political economy of urban educational reform*. New York, NY: Teachers College Press.

Burbules, N. C. (1999). Book review on "Understanding Equal Educational Opportunity". *Teachers College Record, 100*, 882-899.

Cabrera, A. F., & La Nasa, S. M. (2001). On the path to college: Three critical tasks facing America's disadvantaged. *Research in Higher Education, 42*, 119-149.

Choy, S. P., Horn, L. J., Nuñez, A., & Chen, X. (2000). Transition to college: What helps at-risk students and students whose parents did not attend college. In A. F. Cabrera & S. M. La Nasa (Eds.), *Understanding the college choice of disadvantaged students* (pp. 45-63). San Francisco, CA: Jossey-Bass.

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, S95-S120.

DeYoung, A. J. (1987). The status of American rural education research: An integrated review and commentary. *Review of Educational Research, 57*, 123-148.

Gamoran, A. (1987). Stratification of high school learning opportunities. *Sociology of Education, 60*, 135-155.

Hannaway, J., & Talbert, J. E. (1993). Bringing context into effective schools research: Urban-suburban differences. *Educational Administration Quarterly, 29*, 164-186.

Hebel, S. (1999, November 26). *A.P. courses are new target in struggle over access to college in California*. Chronicle of Higher Education, A32.

Hearn, J. C. (1992). Emerging variations in postsecondary attendance patterns: An investigation of part-time, delayed, and non-degree enrollment. *Research in Higher Education, 33*, 657-687.

Heller, D. (Ed.) (2001). *The states and public higher education policy: Affordability, access, and accountability*. Baltimore, MD: Johns Hopkins University Press.

Hossler, D., & Gallagher, K. S. (1987). Studying student college choice: A three-phase model and the implications for policymakers. *College and University, 62*, 207-221.

Howe, K. R. (1997). *Understanding equal educational opportunity: Social justice, democracy, and schooling.* New York: Teachers College Press.

Huang, G., Salvucci, S., Peng, S., & Owings, J. (1996). *National Education Longitudinal Study of 1988 (NELS: 88) research framework and issues.* Washington, DC: US Department of Education, Office of Educational Research and Improvement.

Maeroff, G. I., Callan, P.M., & Usdan, M. D. (eds.) (2001). *The learning connection: New partnerships between schools and colleges.* New York: Teachers College Press.

McDonough, P. (1997). *Choosing college: How social class and schools structure opportunity.* Albany, NY: SUNY Press.

McPherson, M. S., & Schapiro, M. O. (1991). *Keeping colleges affordable.* Washington, DC: Brookings Institution.

National Center for Public Policy and Higher Education (2001). *Measuring up 2000.* San Jose, CA: Author.

Orfield, G. (1988). Exclusion of the majority: Shrinking college access and public policy in metropolitan Los Angeles. *Urban Review, 20* (3), 147-183.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

Perna, L. W. (2000). Differences in the decision to attend college among African Americans, Hispanics, and Whites. *Journal of Higher Education, 71*, 117-141.

Roscigno, V. J., & Crowley, M. L. (2001). Rurality, institutional disadvantage, and achievement/attainment. *Rural Sociology, 66*, 268-293.

Rumberger, R. W., & Thomas, S. L. (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education, 73*, 39-67.

St. John, E. P., Asker, E., & Hu, S. (2001). The role of finances in student choice: A review of theory and research. In M. B. Paulsen & J. C. Smart (Eds.), *The finance of higher education: Theory, research, policy, and practice* (pp. 419-438). New York: Agathon Press, A Division of Algora Publishing.

Shouse, R. C. (1998). Restructuring's impact on student achievement: Contrasts by school urbanicity. *Educational Administration Quarterly, 34*, 677-699.

Smith, M. H., Beaulieu, L. J., & Seraphine, A. (1995). Social capital, place of residence, and college attendance. *Rural Sociology, 60*, 363-380.

Stampen, J. O., & Hansen, W. L. (1999). Improving higher education access and persistence: New direction from a "systems" perspective. *Educational Evaluation and Policy Analysis, 21*, 417-426.

Terenzini, P. T., Cabrera, A. F., & Bernal, E. M. (2001). *Swimming against the tide: The poor in American higher education.* New York: The College Board.

About the Author

Shouping Hu, Ph.D.
Assistant Professor of Education
Seton Hall University
Phone: 973-275-2324
Fax: 973-761-7642
Email: hushoupi@shu.edu
Web: http://pirate.shu.edu/~hushoupi

Shouping Hu is an Assistant Professor of Higher Education in the Dept. of Education Leadership, Management, and Policy at Seton Hall University. His current research concentrates on postsecondary equity and college student experiences and learning. He has taught graduate-level courses in Higher Education Finance, Education Policy Analysis, The American College Student, and Educational Research Design at Seton Hall.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)

Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

This article has been retrieved 7519 times since May 8, 2003

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Education Policy Analysis Archives

Volume 11 Number 15 **May 8, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Teaching Children to Read: The Fragile Link Between Science and Federal Education Policy

Gregory Camilli
Sadako Vargas
Michele Yurecko

National Institute for Early Education Research
and
Rutgers University

Citation: Camilli, G., Vargas, S., and Yurecko, M. (May 8, 2003). *Teaching Children to Read*: The fragile link between science and federal education policy. *Education Policy Analysis Archives, 11*(15). Retrieved [date] from http://epaa.asu.edu/epaa/v11n15/.

Abstract

Teaching Children to Read (TCR) has stirred much controversy among reading experts regarding the efficacy of phonics instruction. This report, which was conducted by the National Reading Panel (NRP), has also played an important role in subsequent federal policy regarding reading instruction. Using *meta-analysis*, the NRP found that systematic phonics instruction was more effective than alternatives in teaching children to read. In the present study, the findings and procedures leading to TCR were examined. We concluded that the methodology and procedures in TCR were not adequate for synthesizing the research literature on phonics instruction. Moreover, we estimated a smaller though still substantial effect (d = .24) for systematic phonics, but we also found an effect for systematic language activities (d = .29) and tutoring (d = .40). Systematic phonics instruction when combined with language activities and individual tutoring may *triple* the effect of phonics alone. As federal policies are formulated around early literacy curricula and instruction, these findings indicate that phonics, as one aspect of the complex reading process, should not be over-emphasized.

Executive Summary

In 1997 the U.S. Congress directed the Director of the National Institute of Child Health and Human Development (NICHD), in consultation with the Secretary of Education, to establish a national panel on research in early reading development. The panel, now known as the National Reading Panel (NRP), was charged with conducting a thorough study of the research, determining what research findings were suitable for classroom application, and recommending methods of dissemination. Six areas of reading were eventually examined, and an influential report was released in December 2000. This report, *Teaching Children to Read*, has stirred much controversy among reading experts, and both critics and supporters have been highly visible in national-level venues. Without question, the report has played an important role in subsequent federal policy regarding reading instruction.

One of the six areas of reading research examined by the NRP was phonics instruction. According to the NRP:

> An essential part of the process for beginners involves learning the alphabetic system, that is, letter-sound correspondences and spelling patterns, and learning how to apply this knowledge in their reading. Systematic phonics instruction is a way of teaching reading that stresses the acquisition of letter-sound correspondences and their use to read and spell words…. (NRP, 2000b, p. 2-89).

Using a research methodology known as *meta-analysis*, the NRP identified 38 experimental and quasi-experimental—meaning a reasonably close approximation to experimental—research studies on phonics instruction. (A meta-analysis can be thought of as a quantitative literature review.) Based on a statistical "averaging" of the outcomes from these 38 studies, the NRP concluded that their findings "provided solid support" for the conclusion that systematic phonics instruction is more effective than alternatives in teaching children to read. Altogether, eleven conclusions were offered regarding the efficacy of phonics instruction, but the above finding is of prime importance.

In their deliberations on research findings, the NRP clearly recognized the ultimate need for instructional decisions to be based on the best empirical evidence and methods of analysis. The NRP recounted that one theme "expressed repeatedly," at a series of five regional public hearings held prior to its work, was the importance of high standards for choosing evidence about what works in reading instruction. The NRP interpreted this to mean that experimental and quasi-experimental studies were most likely to contain reliable, valid, and replicable findings. However, two aspects of the scientific method are important and should be distinguished. The review process, i.e., meta-analysis, is a set of procedures for distilling conclusions and generalizations from research studies. In contrast, the "standards of scientific evidence"—which led the NRP to focus on experimental studies—determine what evidence will be included in the meta-analytic process.

For the purposes of this review, we were primarily concerned with the former aspect, that is, the research review process. Most currently available reviews of the NRP's study have focused on the *interpretation* of the results for phonics instruction while assuming the basic correctness of the measurement and analytic procedures. We did not make such assumptions; rather, we designed an independent study in an attempt to reconstruct the NRP's central findings. As in other types of scientific investigation, replicability is a key criterion for judging the credibility of the NRP meta-analysis, and consequently how seriously we should consider applying its findings.

We began with the *same* 38 studies analyzed by the NRP, but in the course of our analysis, we deleted one study and added three. We then devised alternative plans for extracting and analyzing data from 40 studies (38 – 1 + 3 = 40). Based on these analyses, conclusions were drawn and interpretations made regarding the efficacy of phonics instruction. Though some of the methodological steps taken by the NRP analysts were retraced, our goal was to verify whether an independent team of researchers would arrive at conclusions consistent with those in the NRP report. We did not examine how the original 38 studies were chosen. It would have been useful to examine the full range of the NRP's procedures and findings, including study selection, but this task would have required resources well beyond our means.

In our analyses, we found that programs using systematic phonics instruction outperformed programs using less systematic phonics with $d = .24$. Though this effect is statistically significant, it was substantially smaller than the estimate of the NRP at $d = .41$. (Roughly speaking, $d = 0$ means no effect; $d = .5$ is moderate; and $d = 1.0$ is large.) The systematic phonics effect, moreover, was smaller than the effect for individual tutoring ($d = .40$). Students receiving tutoring had one-to-one instruction as opposed to instruction in small groups or classes. We also found that students who received systematic language activities did better ($d = .29$). This effect is comparable to that of systematic phonics instruction. In addition, standardized tests tended to give larger effects than locally developed instruments ($d = .19$). Overall, we concluded that there is reason to believe that these effects are additive. Systematic phonics instruction when combined with language activities and individual tutoring may *triple* the effect of phonics alone.

Though language activities were included in over 30% of the treatment conditions in the 38 studies, the NRP analysts missed the language effect for one simple reason: they didn't look for it. In our opinion, an approach that recognizes the complexity of reading instruction has the potential to improve the estimates of average effect sizes in *all* substantive areas that the NRP examined including: phonemic awareness instruction; fluency; comprehension; vocabulary instruction; text comprehension instruction; teacher preparation and comprehension; strategies instruction; teacher education and reading instruction; and computer technology and reading instruction. To obtain more accurate estimates of the full range of variables that influence reading, analyses would also benefit from, and indeed may require, a substantially larger sample of studies. In this effort, researchers with substantive, methodological, and classroom experience—as well as time and resources—are necessary to find studies, and to propose and test alternative design strategies. While we applaud the NRP for taking the challenging and difficult first steps in summarizing the extant knowledge on reading instruction, it is clear that substantial resources will be required for completing this essential work.

If the NRP results are taken to mean that effective instruction in reading should focus on phonics to the exclusion of other curricular activities, instructional policies are likely to be misdirected. This interpretation of the data results from a design in which simultaneous influences on reading interventions were not adequately coded and analyzed. In particular, early literacy policies are a timely concern, especially as they are interpreted and applied in the federal Early Reading First Program. Program administrators and teachers need to understand that while scientifically-based reading research supports the role of phonics instruction, it also supports a strong language approach that provides individualized instruction. As federal policies are formulated around early literacy curricula and instruction, it is important not to over-emphasize one aspect of a complex process. Fletcher and Lyon (1998) wrote "a targeted skill cannot be learned without opportunities for practice and application." With this common sense observation in mind, it is not surprising that the research shows a balance of systematic phonics, tutoring, and language activities is best for teaching children to read.

Introduction

In 1997 the U.S. Congress directed the Director of the National Institute of Child Health and Human Development (NICHD), in consultation with the Secretary of Education, to establish a national panel on research in early reading development. The panel, which is now known as the National Reading Panel (NRP), was charged with conducting a thorough study of the research, determining what research findings were suitable for classroom application, and recommending methods of dissemination. Five areas of reading were eventually examined, and an influential report was released in December 2000. This report (NRP, 2000a), *Teaching Children to Read* (Note 1), has stirred much controversy among reading experts, and both critics and supporters have been highly visible in national-level venues (e.g., Manzo, 1998; Pressley & Allington, 1999; Yatvin, 2000; Krashen, 2000, 2001; Garan, 2001, 2002; Ehri & Stahl, 2001; Shanahan, 2001; Coles, 2003). In any case, the report has played an important role in subsequent federal policy regarding reading instruction (Manzo, 2002; Manzo & Hoff, 2003).

One of the five areas of reading research examined by the NRP was phonics instruction. According to the NRP:

> An essential part of the process for beginners involves learning the alphabetic system, that is, letter-sound correspondences and spelling patterns, and learning how to apply this knowledge in their reading. Systematic phonics instruction is a way of teaching reading that stresses the acquisition of letter-sound correspondences and their use to read and spell words…. (NRP, 2000b, p. 2-89).

Using a research methodology known as *meta-analysis*, the NRP identified 38 experimental and quasi-experimental (meaning a reasonably close approximation to experimental) research studies on phonics instruction. Based on a statistical analysis of the quantitative results from these 38 studies, the NRP concluded that:

> Findings [from the meta-analysis] provided solid support for the conclusion that systematic phonics instruction makes a more significant contribution to children's growth in reading than do alternative programs providing unsystematic or no phonics instruction. (NRP, 2000b, p. 2-132)

Altogether, eleven conclusions were offered regarding the efficacy of phonics instruction, but the above finding is of prime importance.

In their deliberations on research findings, the NRP clearly recognized the ultimate need for instructional decisions to be based on the best empirical evidence and methods of analysis. At a series of five regional public hearings held prior to its work, the NRP recounted that one theme "expressed repeatedly" was

> The importance of applying the highest standards of scientific evidence to the research review process so that conclusions and determinations are based on findings obtained from experimental studies characterized by methodological rigor with demonstrated reliability, validity, replicability, and applicability. (NRP, 2000a, p. 1-2)

Two aspects of the scientific method should be distinguished in this desideratum: the "research review process," and the "standards of scientific evidence" that led the NRP to focus on experimental studies.

In this document, we are primarily concerned with the former aspect, that is, the research review process. Most currently available reviews of the NRP's study have focused on the *interpretation* of the results for phonics instruction while assuming the basic correctness of the measurement and analytic procedures. We did not make such assumptions; rather, we designed an independent study in an attempt to reconstruct the NRP's central findings. As in other types of scientific investigation, replicability is a key criterion for judging the credibility of the NRP meta-analysis, and consequently how seriously we should consider applying its findings.

We began with the *same* 38 studies analyzed by the NRP, but in the course of our analysis, we deleted one study and added three (Note 2) others originally identified by the NRP. We then devised alternative plans for extracting and analyzing data from the 40 studies (38 – 1 + 3 = 40). Based on these analyses, conclusions were drawn and interpretations made about the efficacy of phonics instruction. Though some of the methodological steps taken by the NRP analysts were retraced, our goal was to verify whether an independent team of researchers would arrive at conclusions consistent with those in the NRP report. We did not examine how the original 38 studies were chosen. It would have been useful to examine the full range of the NRP's procedures and findings, including study selection, but this task would have required resources well beyond our means.

Our investigation resulted in several major findings. We obtained a statistically significant effect for systematic phonics instruction, but one that was substantially smaller than that of the NRP. Relative to systematic phonics, we also found that individualized instruction (i.e., tutoring v. small group or class) had a substantially larger effect while language-based *instructional activities* yielded a comparable effect. Finally, we concluded that there is no reason to believe that these effects are mutually exclusive. Systematic phonics instruction when combined with language activities and individual tutoring appears to have a much larger effect than phonics alone.

The remainder of this report consists of seven sections:

I. Introduction to Meta-Analysis. A brief introduction to meta-analysis is given.
II. Findings of NRP Study. An overview of the NRP findings on phonics instruction is given along with select results.
III. Reanalysis: Research Questions and Methods. Questions examined by the current study are listed, and methodological issues are described.
IV. Re-Analysis: Results. Quantitative results of the present study are given.
V. Re-analysis: Discussion. The size of the phonics effect is evaluated using results from other meta-analyses and the moderator effects estimated in the present study.
VI. Meta-analysis and Public Policy. Meta-analysis is discussed as a method for resolving controversial issues.
VII. Conclusions. Conclusions and recommendations are given with respect to integrating research, especially with respect to phonics instruction.

I. Introduction to Meta-Analysis

Meta-analysis is a *public analysis* of research findings. It uses publicly available data sources and reveals explicitly to stakeholders how data are selected and analyzed. Private knowledge of data or methodology plays no role. Cooper and Hedges (1994) summarized more elegantly:

> Two decades ago the actual mechanics of integrating research usually involved covert, intuitive processes taking place in the head of the synthesist. Meta-analysis made these processes public and based them on shared, statistical assumptions (however well these assumptions were met). (p. 11)

Nearly a quarter century ago, meta-analysis was developed as a set of statistical procedures for combining the results of many primary studies on a single topic (Glass, McGaw, & Smith, 1981). Previously, there was no effective way to solve the dilemmas of conflicting individual or primary studies. With meta-analysis, each study contributes information in a systematic way, and differences are resolved through statistical analysis.

In a nutshell, meta-analysis is a method of statistically summarizing *quantitative* outcomes across many research studies. Cooper and Hedges (1994) described this method as consisting of five steps:

1. *Problem formulation*. Researchers decide whether a sufficient number of studies exists for a subject of theoretical (e.g., speed of recall) or practical (e.g., class size) interest. These studies usually investigate treatments or interventions in the framework of a comparative research design. (Note 3) For example, we might ask whether students do better in a smaller class (experimental group) rather than a larger class (control group). This step also involves defining a population of interest (e.g., 4th graders) as well as measurements or outcomes (e.g., performance on multi-step math problems).
2. *Data collection: Searching the literature*. Ideally, all relevant studies would be obtained for a meta-analysis. To obtain the most exhaustive sample of studies possible, the researchers must sort through all appropriate reference systems and publications. Additional studies are frequently added by combing through the references of obtained studies as well as databases of unpublished studies. The key idea here is that if a sample of studies is obtained, that sample must fairly represent the entire population of studies to avoid bias (in the same way that the U.S. Census must ensure that hard-to-reach subpopulations are fairly represented).
3. *Data evaluation: Coding the literature*. Trained researchers must extract information about each study's results. A standard list of features (e.g., size of the treatment groups) is developed prior to reading through the studies, even though some of this information may not be reported in many studies. Different researchers who record study information work with common variable definitions so that the information is reliable and comparable across studies. (Note 4) The determination of *what counts as relevant information for coding purposes* should be made by experts who have a thorough understanding of the treatments, populations, and measurements in question. Meta-analysis requires a quantitative measure of effect or outcome, but studies using conceptually similar measures often do not use the same nominal instruments or tests. Therefore, to be able to combine quantitative treatment-control differences across instruments, they must be translated to a common scale. For example, if one wanted to add two measurements, one in centimeters and one in inches, it would be necessary to convert inches to centimeters (or vice versa). This is what an *effect size* (labeled as *d*) ideally accomplishes. It is a translation of the measured effects from different studies into comparable units (in this case, standard deviations). More description is given in Section V on the effect size measure *d*, but as a rule average effect sizes in instructional research tend to range from 0 to about ±1.
4. *Analysis and interpretation*. A central question for all comparative studies is the degree to which the experimental group (sometimes called the treatment group) outperformed the control group. Once effect sizes are computed, statistical analyses are used to estimate the average *d* and its margin of error. Analyses also determine whether certain study features like the duration of treatment influence the effect size. Note that *estimation* of an effect is a different activity than its *interpretation*. The meaning of a measurement in centimeters can be quite different

depending on, for instance, whether we are talking about the following distance of automobiles on a highway or the width of a contact lens.

5. *Public presentation.* At every stage of the meta-analysis, records should be kept regarding procedures. In reporting a meta-analysis, researchers must provide not just statistical results, but also an account of decisions that led to those results. In addition, the meta-analysis is not over until the results are linked to the research issues specified in the first step. In short, the findings must be interpreted and communicated. They must also be qualified, that is, the researchers help readers to understand limitations of the meta-analysis.

While the principles of meta-analysis are scientific, the methods it employs are not purely formulaic. Human judgment is a key element in each of the five steps. In particular, meta-analysts rely on expert judgment for converting narrative descriptions of a study's treatments and subject populations to quantitative measurements. Such coding often requires substantive expertise in addition to research and quantitative skills. (Note 5)

II. Findings of NRP Study

The subgroup of the NRP for Phonics Instruction described the five steps of its meta-analysis in Chapter 3, Part II of *Teaching Children to Read*. In particular, 11 major conclusions were listed (NRP, p. 2-132 to 2-136). The report is well-summarized by Ehri et al. (2001, abstract):

> A quantitative meta-analysis evaluating the effects of systematic phonics instruction compared to unsystematic or no phonics instruction on learning to read was conducted using 66 treatment-control comparisons derived from 38 experiments. The overall effect of phonics instruction on reading was moderate, $d = 0.41$. Effects persisted after instruction ended. Effects were larger when phonics instruction began early ($d = 0.55$) than after first grade ($d = 0.27$). Phonics benefited decoding, word reading, text comprehension, and spelling in many readers. Phonics helped low and middle SES readers, younger students at risk for reading disability (RD), and older students with RD, but it did not help low achieving readers that included students with cognitive limitations. Synthetic phonics and larger-unit systematic phonics programs produced a similar advantage in reading. Delivering instruction to small groups and classes was not less effective than tutoring. Systematic phonics instruction helped children learn to read better than all forms of control group instruction, including whole language. In sum, systematic phonics instruction proved effective and should be implemented as part of literacy programs to teach beginning reading as well as to prevent and remediate reading difficulties.

For additional detail with regard to the overall results, we give the complete text of the first conclusion from the NRP report:

> Children's reading was measured at the end of training if it lasted less than a year or at the end of the first school year of instruction. The mean overall effect size produced by phonics instruction was significant and moderate in size ($d = 0.44$). Findings provided solid support for the conclusion that systematic phonics instruction makes a more significant contribution to children's growth in reading than do alternative programs providing unsystematic or no phonics instruction. (NRP, 2000b, p. 2-132).

Data analyses supporting these conclusions were based on a straightforward design: treatment groups receiving systematic phonics were compared to control groups receiving unsystematic or no phonics instruction. Yet both the experimental and control groups might receive *mixtures* of phonics, language instruction, and other activities. The NRP did examine whether the effect of phonics instruction was influenced by moderator variables, such as socio-economic status or phonics programs. However, no attempt was made to classify the degree of phonics or the mixtures of phonics and other language activities in the groups being studied.

Treatment and Control Group Definitions

In order to understand the overall effect ($d = .41/.44$), it is necessary to understand the characteristics of the treatment and control groups (Note 6). The NRP described treatment groups as including systematic phonics instruction while control groups, though they may have had some phonics instruction, as having various other types of instruction (NRP, 2000b, p. 2-103) with *less* systematic phonics. Thus, the effect size generally signifies the advantage of *more versus less* systematic phonics instruction:

> Whereas some groups were true "no-phonics" controls, other groups received some phonics instruction. It may be that, instead of examining the difference between phonics instruction and no phonics instruction, a substantial number of studies actually compared more systematic phonics instruction to less phonics instruction. (NRP, 2000b, p. 2-124)

Because almost all children received some instruction in phonics during the course of comparative studies, this formulation is realistic. However, the degree of phonics instruction varied from study to study, and it is possible that a treatment in one study could resemble a control in another.

While we believe that the effect size can be a useful measure in such situations, it must be realized that any ambiguity in how comparisons vary across studies adds some ambiguity to the interpretation of the overall or average effect size. The NRP surmised that the effect of such treatment-control variability might be to *underestimate* effect sizes. In many cases, however, children receiving systematic phonics instruction were also receiving activities consistent with the aims and purposes of whole language. Thus, uncontrolled mixtures might also serve to *overestimate* the effects of phonics instruction.

Others have written about the false dichotomy between language and phonics instruction (e.g., Fletcher and Lyon, 1998). (Note 7) A number of phonics instruction treatments are described in the NRP report including synthetic, analytic, analogy, onset-rime, phonics through spelling (NRP, 2000b, p. 2-99), and embedded phonics. Many contain some degree of language instruction. For example, although "embedded" phonics was not defined in the NRP report, Foorman, Francis, Fletcher, Schatschneider, and Mehta (1998) described their "embedded code" treatment as including "whole-class activities such as shared writing, shared reading, choral or echo reading, and guided reading" (p. 40). In addition the teachers would "frame a word containing the target spelling pattern during a literacy activity" (p. 40). Consequently, the treatment is consistent in some important respects with language-based instruction, though it can also be described as a type of phonics instruction. While such treatments defy simple labels, they can be coded on various dimensions that more accurately describe the "package" of treatment conditions. Analyses can then be undertaken to sort out the unique effects of various instructional activities and conditions.

Outcome Variables and Units of Analysis

The NRP subgroup on phonics instruction computed effects sizes for dependent variables that fit into one of 7 categories (also see Table 1) (Note 8):

1. Word ID
2. Decoding
3. Spelling
4. Comprehension
5. Nonword reading
6. Oral reading
7. General reading

Table 1
Dependent Variable Categories.

Category	Label	NRP Label
1	decoding regular words	decoding
2	decoding nonwords	nonwords
3	sight word ID	word ID
4	spelling	spelling
5	comprehension	comprehension
6	oral reading	oral reading
7	general reading	general reading
8	language	*
9	phonemic awareness	*
10	alphabetic knowledge	*
11	vocabulary	*
12	writing	*
*Category not used in NRP study		

For each category within each treatment-control comparison, it is our understanding (NRP, 2000a, p. 1-10) that either mean or median effect sizes were computed for each cohort of students when results for more than one test instrument were available. In some cases, studies did not report measures for some categories, in which case the category was left blank (i.e., a "missing value") in Appendix G. At most, one effect size was reported for each category for each cohort/comparison.

Importantly, measures were excluded from this classification if they were used during (or as part of) phonics instruction (NRP, 2000b, p. 2-110). Such effect sizes would be expected to be larger due to "teaching to the test." No distinction was made between standardized and experimenter-devised tests. Because standardized tests are targeted to a wider range of ability, the NRP surmised that they might be less sensitive to change and thus "underestimate effect sizes slightly" (NRP, 2000b, p. 2-111).

Criticisms of the NRP Meta-Analysis

Three prominent criticisms of the NRP meta-analysis of phonics instruction have spurred public debate. The first concerns methodology; the second concerns the link between evidence and conclusions; and the third, the procedures with which research activities were conducted.

The first criticism is that a narrow population of children was represented in the 38 studies that comprised the meta-analysis (Garan, 2002). In particular, Garan argued that many of the studies did not include "normal readers" and none included groups of advanced readers. Thus, it would be difficult to generalize the findings broadly across typical populations of students. The second criticism is that the term "reading" was not used in a consistent manner; the term reading can refer to simple "word calling" (e.g., a response to the question "Can you say this word?"), but it can also refer to the ability to derive meaning from connected text (Yatvin, 2002). If it is said that "Phonics instruction improves reading," it is important to know what kind of reading is signified. The third criticism was that the process used to conduct and report the meta-analysis was flawed. According to Yatvin (2002), the NRP study on phonics instruction was completed in a very short time. In October, 1999, five months before the due date, a determination was made that the completion of the study required resources beyond the capacity of panel members, and it appears that a researcher who was not a member of the NRP was commissioned to conduct the meta-analysis. (Note 9) Upon completion of the study, again due to time constraints, the panel originally in charge of designing and conceptualizing the research had only four days to review the final report before it went to press. Yatvin also observed that only one panel member (Yatvin) had teaching experience, and thus the NRP had little expertise for the purpose of linking research findings to practice.

The NRP addressed some of these issues. The 38 studies provided 66 (Note 10) treatment-control comparisons, and of these, 23 comparisons included normal readers (about 35%). In regard to the second criticism, the NRP found that:

> The majority (76%) of the effect sizes involved reading or spelling single words while 24% involved reading text. The imbalance favoring single words is not surprising given that the focus of phonics instruction is on improving children's ability to read and spell words. (NRP, 2000b, p. 2-92)

Even from this brief quote, it is clear that a necessary distinction must be made between "word reading" and conceptualizations of reading that imply understanding of connected text. "Word reading" is just one connotation of reading, yet the distinction isn't maintained consistently in formal documents. For example, in Ehri and Stahl's (2001) rebuttal to Garan (2001) (Note 11) they reported that clear evidence was found to support the conclusion that

> Systematic phonics instruction was found to be more effective than unsystematic phonics instruction or no phonics instruction in helping students learn to *read* [emphasis added]. (Ehri and Stahl, 2001, p. 18)

One could define reading as "reads single words in isolation," which would be consistent with the NRP's data analyses. But reading could also be defined as "reads connected text," that is, sentences or stories. Obviously, one's sense of the study's outcome—as represented in the above quote—depends almost entirely on how reading is defined.

The third criticism was that not enough time was allotted to carry out the charge of Congress, and that the final report was not subjected to formal review. In fact, the study was under intense time pressure from inception. According to Yatvin, who wrote a minority addendum to the final report,

> In fairness to the Panel, it must be recognized that the charge from Congress was too demanding to be accomplished by a small body of unpaid volunteers, working part time, without staff support, over a period of a year and a half. (The time Congress originally allotted was only 6 months.) (Yatvin, 2000, p. 2)

Whether the resources and time were sufficient to carry out such an important study is now a moot issue. The question of interest is whether the meta-analysis conducted by the NRP is sufficiently reliable and valid for guiding instructional policy in early reading. In the present study we address the topic of whether the central NRP results can be replicated by a different team of analysts. A successful replication would provide convincing evidence of accuracy and allay concerns about study logistics.

III. Reanalysis: Research Questions and Methods

The NRP results were given for 11 central questions regarding phonics instruction. In this re-analysis, we will be concerned primarily with two of these: "Does systematic phonics instruction help children to learn to read more effectively than nonsystematic phonics instruction or instruction teaching no phonics?" (NRP, 2000b, p. 2-132); and "Is phonics instruction more effective when it is introduced to students not yet reading, in kindergarten or 1st grade, than when it is introduced in grades above 1st after students have already begun to read?" (NRP, 2000b, p. 2-133).

Using public—that is, published—accounts of data and methodology, we re-examined the evidence offered by the NRP on the efficacy of phonics instruction. We designed an effect size database, recomputed effect sizes for all outcomes available, and then carried out analyses in which effect sizes were related to study characteristics. One study by Vickery, Reynolds, and Cochran (1987), which is described in Appendix C, examined the effect of the *same* treatment on remedial and nonremedial students. Because there was no control group, we deleted this study from our database (see inclusion criteria on p. 2-108 to 109 in NRP, 2000b). We included another three studies that were identified by the NRP but not included in their meta-analysis. These are described in Appendix A of this report. Thus, our database was constructed from 40 studies originally identified by the NRP; however, the merits of the original NRP sample or sample selection process is beyond the scope of the present study. (Note 12; Note 13)

Our analytic strategy had several components. We selected a unit of analysis, defined alternative weighting schemes, and used multiple regression to identify the unique contributions of variables that moderate the treatment. By *moderator* variable, we mean a component of treatment delivery that leads to a stronger or weaker effect. Four new moderator variables were constructed for specifying the treatment conditions: the degree of phonics systematicity; degree of coordinated language activities; whether treatments were regular in-class or pullout programs; and whether basal readers were used. These variables, which were coded from the research studies by means of rubrics, provided the explanatory power missing from the simple comparative design used in the NRP analyses. That is, the NRP design did not fully account for variation in the mixtures and degrees of treatment delivered to both experimental and control groups. Other moderators were borrowed from Appendix G of the NRP report. Using regression analysis, we then predicted treatment outcomes (i.e., effect sizes) with the four new moderators and: the size of the instructional unit (tutoring, small groups, class); whether treatment conditions were randomly assigned; whether standardized tests were used; and the age (Note 14) of students.

There are two important design facets in a meta-analysis. The first is a design for *data collection*, while the second parallels the usual sense of the word in the phrase *experimental design*. That is, there is one design for data collection, and another for analysis. In order to address the weaknesses of the simple comparative design of the NRP study, we coded moderator variables, but we also planned for a more complete use of the information within each of the 40 studies. In particular, we distinguished untreated control groups from "alternative" treatments, and included both, as described below. This can be likened to filling out the cells of—or balancing—an experimental design, while the increasing the number of studies adds to sample size. The recognition of this distinction is not evident in the NRP analytic plan.

Database Design

Including Groups for Comparison. As noted above, in each study the NRP designated as the control a group with less systematic phonics than the treatment group (or groups). Ironically, this procedure in some cases led to ignoring information from groups labeled as "control" by the authors of the primary studies. For example, in the study by Lovett, Ransby, Hardwick, Johns, and Donaldson (1989) three groups were used: the Decoding Skills Program (DS), the Oral and Written Language Stimulation program (OWLS), and a Classroom Survival Skills program (CSS). The third group was described as a "control procedure in which subjects received the same amount of clinic time and professional attention as those in the experimental remedial programs" (p. 96); however, CSS students received training in activities that didn't include reading. It appears that non-treatment controls such as the CSS group were excluded from the NRP study when programmatic controls like OWLS were present. Thus, the NRP effect sizes for Lovett et al. (1989) are based solely on the comparison of the DS to the OWLS program.

In such cases, we computed effect sizes for DS versus CSS *and* OWLS versus CSS. However, we coded (with treatment indicators determined by rubric codings) the DS program as having systematic phonics instruction while the OWLS program was coded as language-based. This strategy yields an important source of information for disentangling treatment effects because untreated control groups can provide a common basis for comparison across studies. The component effects of treatment mixtures may then be more accurately identified.

Defining Control Groups. More than one control group may have been available for computing effect size. For example, in Foorman et al. (1998), there were four groups described as: direct code (DC), embedded code (EC), implicit code-research (IC-R), and implicit code-standard (IC-S). It appears that the NRP analysts used the IC-S group as the control even though the authors of the study asserted that comparisons among IC-R, DC, and EC provided the most relevant information about instructional differences because the IC-R group controlled for teacher training.

We decided to use the IC-R group for computing effect sizes based on the general rationale in this paragraph, which we used for all studies. The most valid control was taken as the group that received the same kinds of treatment activities (e.g., individual attention, duration of treatment), but not the treatment itself—either language or phonics. This would serve to control for as many background variables and moderators as possible. For instance, if there were a choice of control between two groups that did not involve phonics or language instruction, then we would use this rule to choose the control. We coded systematic language programs as treatments unless there was not another control group available. In a study with only a phonics group and a language group, we compared the phonics to the language group to obtain the effect size, but coded the comparison as being Phonics v. Language rather than Phonics v. Control. At least three possible classes of comparison (phonics-control, language-control, and phonics-language) were defined by the rubric indicators.

In summary, we included control groups having no systematic phonics or language interventions, whereas the NRP analysts did not. However, when two control groups were available, we chose the one most like the treatment group in terms of characteristics ancillary to the intervention.

Coding Rubrics and Inter-Rater Reliability

We coded the characteristics of both treatment and control groups with rubric indicators. The rationale for this practice is that coding is a measurement process, requiring inference, and not a simple reading of a study. Since coding is a measurement process, its scientific warrant should be established by demonstrating inter-rater agreement. The credibility of the limited moderators coded by the NRP team was also established by demonstrating high inter-rater agreement.

In Table 2, the rubrics are given that were used to code treatment characteristics. We distinguished among three levels of phonics instruction; two levels of language; basal reader usage; and supplemental/pullout versus regular in-class instruction. Rubric codings provide a richer quantitative description of studies in which instruction is comprised of mixtures of phonics, language, and other elements. For each study, three independent codings were obtained. The first codings were given by the authors of the present study, each of whom had participated in all aspects of at least one previous meta-analysis. None had previously participated in a study of phonics or whole language instruction, and none had taken a public position in the phonics versus whole language debates. The second and third codings were provided, respectively, by an experienced reading teacher and a university professor, each with a national reputation in reading instruction.

Table 2
Rubrics for Coding Treatment Conditions.

Phonics	
ng	No information in study to infer code.
0	No specific phonics intervention was given. In most cases, we know that it is highly probable that students received some kind of phonics activity, especially for longer interventions. Moreover, even if no phonics instruction was associated with the treatment delivered, it may have been the case that other instructional activities (external to the treatment) included phonics. In short, we were not able to distinguish among these possibilities.
1	Treatment specifically included phonics activities, but treatment activities were not described in detail as being direct, systematic instruction. Organized phonics were embedded in language instruction.
2	Treatment was described as including direct, systematic phonics instruction. It was most often the case that this description specifically included blending.
Replace	
ng	No information in study to infer code.
0	Treatment did not replace regular classroom instruction. In some cases, the treatment consisted of a supplemental program. For example, students received treatment at facilities outside of schools (e.g., hospital setting on Saturdays).
1	Treatment was regular classroom instruction, or the treatment completely replaced regular classroom instruction.
Basal	
ng	No information in study to infer code.
0	Basal reader was not used.
1	Treatment was described as including a basal reader, or it was highly probable that a basal reader was used. For example, a 4-year treatment consisting of regular classroom instruction almost certainly used a basal reader at some point, even if it was not specifically mentioned.
Language	
ng	No information in study to infer code.
0	No systematic or formal language activities were included.
1	Language-based (non-basal) treatment was given. This may have consisted of whole word or whole language programs.

For each effect size computation, both experimental and control groups were coded according to the rubrics, allowing for the possibility that any group could be coded as having both phonics and language instruction. However, no phonics treatment labeled as such ever had less systematic phonics instruction than the group chosen as the control, though both groups may have had language instruction. In some cases, study information for coding a rubric was denoted as "not given" by one or more coders. Our guiding principle on this matter was that evidence of "presence" was required in order to make inferences regarding the effects of a rubric variable. We converted "not given" responses to zeros. For example, if a study did not report that basal readers were used, but it was known that the reading program formally included basal readers (and did during the timeframe of the study), then assuming their presence was a relatively safe inference. However, for less familiar or unknown reading programs, it was safest to assume basal readers were not used. In short, the conservative approach to coding was to require evidence of "presence" rather than "absence" when linking treatment or moderator indicators to study outcomes.

In Tables 3a-3d, agreement analyses are given for each of the four rubric variables separately. Under the column labeled "Judges Codings" the number of each possible combination (i.e., unordered triplet) of three codes, one for each judge, is given. Overall, there was substantial agreement among coders, given the evidence-of-presence requirement. In addition to the data in Tables 3a-3d, it is also useful to consider that three raters operating at random with 95 total comparisons would only have an expected value of about 10-11 matches with a 3-point rubric, and only about 23-24 on a 2-point rubric.

Table 3a
Inter-rater Agreement for the Phonics Rubric
(Cronbach's alpha for this rubric was .95)

Judges Codings	***n* (95 total)**	**Cumulative Percent**	**Agreement Type**
0,0,0	22	23	Perfect
1,1,1	13	37	Perfect
2,2,2	31	69	Perfect
0,0,1	12	82	Adjacent
0,1,1	6	88	Adjacent
1,1,2	4	93	Adjacent
1,2,2	5	98	Adjacent
0,1,2	1	99	———
unclassed	1	100	———

Table 3b
Inter-rater Agreement for the Language Rubric
(Cronbach's alpha for this rubric was .79.)

Codings	***n* (95 total)**	**Cumulative Percent**	**Agreement Type**
0,0,0	55	58	Perfect
1,1,1	11	69	Perfect
0,0,1	9	79	———
0,1,1	19	99	———
unclassed	1	100	———

Table 3c
Inter-rater Agreement for the Basal Reader Rubric
(Cronbach's alpha for this rubric was .82.)

Codings	***n* (95 total)**	**Cumulative Percent**	**Agreement Type**
0,0,0	52	55	Perfect
1,1,!	18	74	Perfect
0,0,1	19	94	———
0,1,1	5	99	———
unclassed	1	100	———

Table 3d
Inter-rater Agreement for the Pullout Rubric
(Cronbach's alpha for this rubric was .87.)

Codings	***n* (94 total)**	**Cumulative Percent**	**Agreement Type**
0,0,0	30	32	Perfect
1,1,1	40	74	Perfect
0,0,1	12	87	———

0,1,1	11	99	———
unclassed	1	100	———

When we encountered a difference among coders, the final code was chosen as the consensus code in almost all cases. In the few cases where a 2-of-3 majority was not obtained, codes were averaged. For example, in one comparison the level of phonics was given codes of 0, 1, and 2, and in this case, the results were averaged resulting in a code of 1.0. Given the "majority rules" principle (Orwin, 1994), the three judges were overruled 24, 25, and 56 times out of 404 coding instances. This translates into overruled percentages of about 6%, 6%, and 14%, respectively. Thus, an individual judge's code was retained in a minimum of 86% of the coding instances. The coefficient alphas were relatively high at .95 (degree of phonics), .79 (presence of language activities), .82 (use of basal readers), and .87 (regular v. pullout program).

The effort in detailing treatment conditions is important for making a strong statistical link between treatments and outcomes, and would ideally be planned in the design of the study and coding protocol. Even so, a dose of realism is required in this effort. The typical study examined lacked clarity with regard to treatment conditions. It was not uncommon that a total of three or four sentences were devoted to describing an intervention. Though we feel 69% is an acceptable rate of *perfect* agreement for level of phonics instruction (with a reliability of $\alpha = .95$, see Table 3a), at least part of the 31% of disagreement can be attributed to the lack of clear descriptions of independent variables. Some "measurement error" reflects ambiguous descriptions rather than ambiguity inherent in the judgment process.

In the analyses reported below we used, but did not code, other moderator variables in addition to the rubric indicators. These moderators were borrowed directly from the NRP study including treatment unit, age/grade, SES, and reading ability. We coded, but did not obtain inter-rater agreements, for several additional variables for each effect size in our database including the size (n) of each treatment/comparison group, whether the effect size was from a randomized study, and whether the effect size was from a standardized instrument.

Dependent Variables

Variable Categories. All effect sizes were recomputed for all available outcome measures that could be considered as falling into one of the categories in Table 1. In some cases, we considered outcomes that the NRP did not use, such as alphabetic knowledge, which refers to how well students can connect phonemes to graphemes. Though these measures fell outside the range of the NRP's definition of reading, we felt the information was useful.

Effect Size Computation. Most criticisms and counter-criticisms of the NRP report accept as their starting point the computed effect sizes as obtained by the NRP analysts. We did not use the published NRP effect sizes because independent computation is more consistent with the goals of a validation study based on the merits of replication. Therefore, one major focus of the present study is computational: Can the general effect size obtained by the NRP analysts be replicated? However, since we recomputed effect sizes based on a different design (than the one used by the NRP) for experimental-control comparisons, a one-to-one comparison was not possible. For this purpose, we devised an approximate method of comparison (described in Section IV).

Computing an effect size can pose a difficulty with which meta-analysts are all too familiar, but one that may not be transparent to a consumer of meta-analytic information. In studies that do not report the necessary information for a simple computation, information must be pieced together—sometimes using specially designed procedures that may require a number of assumptions. In this section we review a number of these issues that are pertinent to the studies on phonics instruction. Although the DSTAT program was used for computations (Johnson, 1989) by the NRP team, it is often the case that judgments must be made as to what information to enter into the program; different choices may yield different results even when calculations are error-free. In a number of instances, the NRP team may not have appreciated the complexities of computing the effect size d, or they did not provide rationales for their methods. In this regard, we provide several clarifications below for facilitating accurate effect size computation.

Although the NRP cites Cooper and Hedges (1994) regarding formulas for computing effect sizes, the basic formula given by the NRP (NRP, 2000a, p. 1-10) and reproduced below is incorrect:

$$(1)\quad \frac{\overline{X}_T - \overline{X}_C}{.5(s_T + s_C)}$$

Compared to the pooled effect size estimator (g) given by Hedges (1985, p. 78)

$$(2)\quad \frac{\overline{X}_T - \overline{X}_C}{\sqrt{\dfrac{v_T s_T^2 + v_C s_C^2}{v_T + v_C}}}$$

we can see that while the numerators of (1) and (2) are the same, the denominators differ (v_t and v_c are the degrees of freedom for the experimental and control groups, respectively). Moreover, it can be shown that the effect size given by (1) is always larger than that given by the standard formula in (2). The magnitude of this difference is not large, however, and the NRP calculations appear to have been performed with the correct formula. Nevertheless, it is important to communicate established procedures in a public document. If a nonstandard formula is used, a justification should appear in text, but we know of no justification for the formula in (1). (Note 15)

For the most part, we computed effect sizes in a manner consistent with general methodological descriptions given in the NRP; the Hedges correction was used in all cases (Hedges & Olkin, 1985, p. 81). However, because few in-depth details were provided (e.g., NRP 2000b, pp. 2-110 to 2-111), we used several additional guidelines for the current study:

> a) Standard deviations were pooled across all posttest treatment and control groups within a cohort of students to create a common denominator. Hedges effect size adjustment was applied to *g* to arrive at *d* (using degrees of freedom based on the pooled sample).
>
> b) When pretest means were available, effect size numerators were computed as differential average gains to help control for pre-existing differences. Effect sizes, according to the first guideline, were then obtained via division with a common posttest standard deviation. If covariance adjusted effects were reported, these were used in the numerator instead of the difference between average gains of treatment and control groups.
>
> c) When testing was carried out on more than two occasions during a treatment intervention, we computed gains based on pretest and immediate posttest means. If a treatment spanned several years (or grades), we computed an effect size for the first year using the second guideline. For each ensuing year separately, we computed an effect size using the previous year's posttest as the following year's pretest.
>
> d) Effect sizes were computed with custom programming developed for each individual study rather than using one of the available software products. For the most part, calculations were based on formulae given by Cooper and Hedges (1994).

Units of Analysis. In some cases, classrooms or even schools are used as the units of analysis rather than individual students, and this phenomenon did occur in the set of phonics instruction studies. In this case, classes (or schools) are the *units* of observation, and class means comprise the data to be analyzed. The formula given in (2) typically pools individual level standard deviations that are first calculated with the formula:

$$(3) \quad s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

When individual observations are group means, however, the estimate of variability, *s'*, is

$$(4) \quad s' = \frac{s}{\sqrt{n}}$$

which is the formula for the standard error of the mean. Upon comparison, it can be seen that (4) will be smaller than (3) depending on *n*, which is the class (or school) size. Therefore, an effect size using class means will be larger than one based on individual student scores by the multiplicative factor $\sqrt{n}$. With moderately sized classes, the use of means can result in substantially larger effect sizes, but *these are not comparable* to the effect sizes of other studies whose units of observation are students. To remedy this disparity, effect sizes must be translated to the individual metric. (Note 16) As we shall see below, the greatest discrepancy (between a recomputed and original NRP effect size) was due to a unit of analysis problem.

Weighting Studies

In the NRP study, effect sizes were computed for each experimental treatment. For example, if there were one outcome variable, two distinct phonics-based treatments (A and B), and one control group (C), two effect sizes would be computed (A versus C, and B versus C). If there were two or more outcome variables in a dependent variable category (e.g., two spelling tests), the effect sizes for A-C and B-C would be averaged separately within this category.

Because the NRP reported 66 comparisons from 38 studies, some studies contributed more than one effect size. For example, one study by Vickery et al., 1987, contributed 8 comparisons—4 grade level cohorts crossed with two levels of remediation. There are a number of methods for computing the overall average *d* in this situation. First, one could compute the simple average across the 66 comparisons given in Appendix G of the NRP report (NRP, 2000b, pp. 2-169 to 2-175), which results in a mean of .46. This is close to the value .41 which was reported by Ehri, Nunes, Stahl, and Willows

(2001). Implicit in this procedure is that the Vickery et al. (1987) study receives 8 times the weight of a study that contributed a single effect size—because it examined 8 distinct treatment-control cohorts. (Note 17)

A second method consists of weighting studies by the total *n* of the comparison (treatment + control); in other words, comparisons with larger *n*s would receive more weight. This was the method used by the NRP, and results in mean $d = .41$. In the NRP study, the rationale was given that

> The subgroups [committees] weighted effect sizes by numbers of subjects in the study of comparison to prevent small studies from overwhelming the effects evident in large studies. (NRP, 2000a, p. 1-10).

With this practice, however, large studies overwhelm small studies. For example, in Gersten, Darch and Gleason (1988), data from 1973-1974 are available for two cohorts of children with a total $n = 242$. Treatments were provided at the class level. In contrast, one comparison described by Gillon and Dodd (1997) contains $n = 10$ students in two groups. Given a simple weighting by *n*, the latter study would have about 1/24th the weight of the former study. It is our opinion, that this weighting practice should not be automatic; application of statistical weights necessarily gives studies using classes more weight than studies using small groups or tutoring.

In a third method for averaging across studies, separate *studies* are given equal weight. In this case, the 8 effect sizes in Vickery et al. (1987) would each receive a weight of 1/8; and the weights would sum to 1.0. This weighting practice would be repeated for all studies resulting in a set of weights that would sum to exactly the number of independent studies. In the NRP study, this weighting procedure results in an mean $d = .54$. (We note that this effect is *larger* than the estimate reported in *Teaching Children to Read*, but stay tuned.) In our opinion, this approach makes sense when a set of effect sizes is relatively homogenous. Though Shadish and Haddock (1994) asserted that "all things being equal," weighting sample size is the most widely accepted practice, Hedges and Olkin (1985) cautioned that statistical weights should be considered only in cases with homogenous effects sizes:

> Before pooling estimates of effect size for a series of *k* studies, it is important to determine whether the studies can reasonably be described as sharing a common effect size. (p. 122)

Thus, "all things being equal" can be accurately interpreted as "sharing a common effect size."

A fourth method of weighting represents a compromise between statistically weighting and equally weighting studies. Let the statistical weights be labeled as *WGT1*, and let the equal representation weights be labeled *WGT2*. A compromise between the two weight types can be achieved by taking *WGT3 = WGT1*WGT2*. In the latter approach, consideration is given both to study representation and sample size. See Table 4 for definitions of the three types of weighting. For the analyses in the present report, we examine regression estimates derived from the weighting systems represented by *WGT1* and *WGT3*.

Table 4
Definitions of Alternative Unit Weights:
Equal Representation, Optimum, and Compromise.

WGT1:	If a single study contributed *k* records to the aggregated database, the *equal representation* weight was defined as: $$WGT1 = \frac{1}{k}$$
WGT2:	For a particular record in the aggregate database, the total number of observations for the treatment and control groups was: $$n_{TOT} = n_T + n_C$$ The weight was then taken as n_{TOT}. Rather than using this approach, we opted to use the *optimum* weight defined by Hedges and Olkin (1985, pp. 86 & 110) as: $$WGT2 = \left(\frac{n_{TOT}}{n_T n_C} + \frac{d^2}{2(n_{TOT} - 2)}\right)^{-1}$$ After examining the distribution of *WGT2*, we set a maximum value so that the highest values were no more than 15 times larger than the smallest values.

WGT3:	Given *WGT1* and *WGT2*, this *compromise* weight *WGT3* was computed as: $$WGT3 = WGT1 * WGT2$$

How studies should be weighted is a critically important issue, because different weighting methods may give different results. Ironically, the NRP choice resulted in large studies effects overwhelming those of smaller studies, and the consequences of using this choice should be carefully considered. In our database of effect sizes, the test of homogeneity was highly significant ($Q = 813.46$, 223 df; equivalent z-statistic is roughly $z = 27.94$) indicating that the studies did *not* share a common effect size (i.e., the hypothesis of homogeneity was rejected). Statistical weights may be inappropriate for the NRP data because of potential qualitative differences between small and large studies. Moreover, some studies included multiple comparisons, and statistical weighting gives such studies many times the influence of studies with a single comparison. A procedure in which studies are given equal weight may provide the most "equitable" reading of the experimental literature, but a compromise, which balances representation and statistical precision, may also be useful. Other procedures may also be defensible, but in any case an explicit justification should be provided.

The issue of weighting involves notions that are fundamental to the ideals of meta-analysis. Though it can be understood as a statistical issue, weighting can also be understood relative to the questions "What counts as evidence?" and "How should evidence be accumulated?" What counts as research in education usually comes in the form of a "study" in which an author analyzes data and reports conclusions based on those analyses. The results from a single study may be extremely trustworthy and valuable, and so a problem arises when we wish to "sum" the evidence in two or more studies. Other things being equal, conclusions from two studies using the same data would not be valued equally to conclusions from two independent studies. Yet the problem is not simply to determine appropriate weights for different studies, but how to understand the role of cumulative evidence vis-à-vis the role of in-depth knowledge flowing from a single, well-executed study.

Meta-analysis is a systematic method for summarizing the knowledge inherent in a research literature. Information concerning study outcomes based on unreported or private knowledge can obviously not add to this summary, even though what is actually learned from a study *does* include unreported and private knowledge. The truth of analytic conclusions can only be linked to "reports" of empirical investigations. This assertion is very different from the "garbage in—garbage out" axiom, which implies that a simple "truth in—truth out" model is possible for synthesizing research studies. The process of establishing warrants for conclusions is different in meta-analysis than it is in primary research since in published primary studies authors have direct access to contextual information (e.g., vested interests) that is not printed, but nonetheless influences reported conclusions. One important assumption of meta-analysis is that the effects of unreported information will "average out" across independent studies. This is why fair representation and appropriate weighting strategies are such important prerequisites to valid conclusions.

Case Studies

The quality of any meta-analysis is fundamentally based on studies that meet inclusion criteria. In the NRP phonics instruction meta-analysis, the foremost criterion was that "Studies had to adopt an experimental or quasi-experimental design with a control group." (NRP, 2000a, p. 2-108). In addition studies had to appear in a refereed journal after 1970, had to provide information for testing the efficacy of phonics instruction on reading, and had to report statistics necessary for computing effect sizes. Having obtained such studies, information was coded and analyses were conducted. The goal of the NRP meta-analysis was to identify reliable and replicable results in the area of early reading.

In Appendix B, we provide a perspective on three studies that met these inclusion criteria. Our goal is to provide readers with a deeper familiarity with the literature, one that extends beyond the typical boundaries of a meta-analysis. This is important for illustrating how well the inclusion criteria performed in obtaining methodologically rigorous studies, and for giving a more salient notion of the confidence with which we can generalize. Indeed, cases studies were also included in the NRP report because of their descriptive value. We emphasize that the studies in Appendix B of the present report are given for the purpose of illustration—issues arise with any study put under a microscope.

The case studies serve to illustrate methodological issues in a number of areas including: choice of control group, unit of analysis, and study selection criteria. While all three studies use quasi-experimental designs, a more in-depth examination of these can facilitate a practical understanding of the variety and limitations in this design approach to reading research. In our judgment, these studies are representative of, if not of higher quality than, the entire set of 40 studies.

IV. Re-Analysis: Results

The NRP used cohort comparisons as the unit of analysis, and then applied statistical weights. (Note 18) We used two strategies for weighting. (Note 19) The first strategy is the meta-analytic equivalent of "one person one vote" representation. In the second compromise strategy, we combined statistical (inverse variance or comparison n) weights with equal representation weights. We remind the reader that the usefulness of weighting—as well as that of the entire meta-analytic enterprise—depends on how well a set of studies represents the research literature.

The process of coding resulted in obtaining 491 effect sizes from 40 studies for 12 dependent variable (DV) categories. The Vickery study (described in Appendix C) was deleted due to lack of a control group. This left 37 original NRP studies, to which we added three studies with phonemic awareness outcomes. It appeared to us that the latter three studies (described in Appendix A), which were identified but not included in the NRP analysis, met the inclusion criteria of the NRP. Each of these studies, which contributed 7 records total to the database, included at least one reading outcome from categories 1-6 in Table 1. This database did not include effect sizes from follow-up comparisons. The data file contained one data record for each effect size. However, single studies often contributed more than one effect size for a DV category. Moreover, a *d* for the same outcome variable category might be computed for more than one cohort within a study. To manage this redundancy, we aggregated effect sizes to the comparison level within each DV category within a study. This resulted in a primary analysis file of 225 observations (out of a possible 480, which equals 40 studies multiplied by 12 DV categories); 60 of these represented DV categories not included in the NRP study. For each case in the aggregated file we included moderator variables such as duration of treatment, size of the treatment unit, rubric codes, and the like.

Our unit of analysis was "comparison." That is, if a study compared one treatment to one control group, and measured two outcomes, then there were 2 effect size records for one comparison. In some cases, a study had two treatment groups (T_1 and T_2) and one control group (C); in this case with two outcomes, there were 4 effect size records (T_1 v. C, and T_2 v. C. crossed with two outcomes). Equal representation weights were then obtained as the inverse of the number of records per study. Multiple cohorts were averaged, if they existed, within comparison unless the treatment conditions changed across time.

Agreement with the NRP Study

Because of the design difference between the NRP meta-analysis and the present reanalysis, it is not possible to compare the effect sizes for the two studies directly. However, if effect sizes are aggregated to the study level (excluding studies with TP = 0), we can examine the consistency of the two sets of effect sizes. In Figure 1, the scatter plot shows that two studies (labeled 12 and 53) appear as outliers. Study 53 contained an (d = 8.79) outlier and appeared to have been removed from most, if not all, NRP calculations. In study 12, effect sizes were computed by the NRP team with class means; the required conversion of the pooled standard deviation to the individual metric was not made. With these two studies removed, r^2 = .754 for effect sizes based on the original 7 NRP categories (e.g., Nonwords, Decoding, etc.) with TP=1 or TP=2.

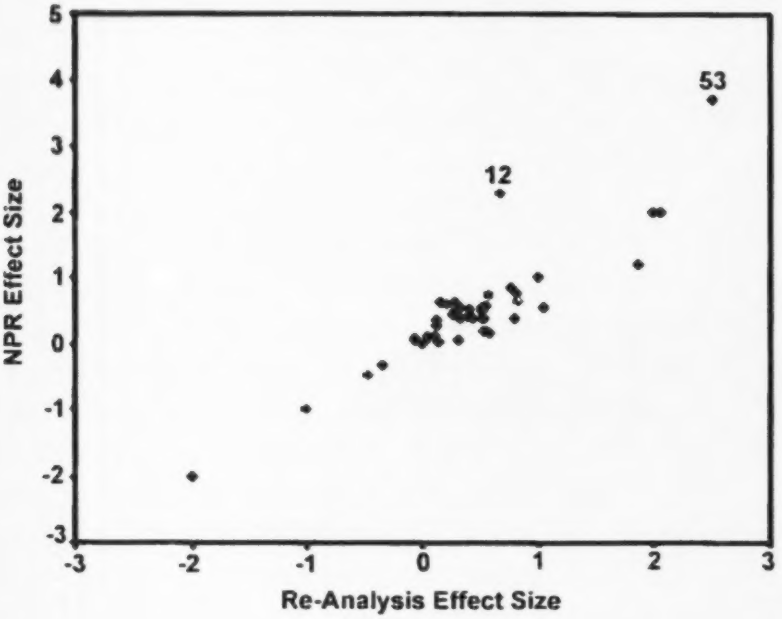


Figure 1. Scatterplot of NRP Calculated Effect Sizes and Our Re-Analysis Calculated Effect Sizes

Again with studies 12 and 53 removed (as well as Vickery et al., 1987), the overall averages of these two sets of effect sizes do not significantly differ using a paired samples t-test (t = -.447, p = .658, 33 df). Clearly, the same general information for effect size was obtained, though a higher level of agreement (correlation) would be desirable. We did not have the disaggregated NRP effect sizes, that is, Appendix G reports effect sizes aggregated by the outcomes classification. For the most part, it was not possible to compare specific effect sizes directly.

Level of Phonics Instruction

We computed the overall average *d* in a different way than the NRP analysts who first computed an average for each cohort, and then computed a weighted average of these (i.e., the cohort averages) across studies. We obtained averages directly from our database, using "equal study representation" and "compromise" weighting. The analysis of central interest is the difference between systematic and less systematic phonics, since the latter is what many, if not most, students already receive. We used the TP rubric variable (scale 0–2) to describe the level of phonics as a break variable for computing the weighted means given in Table 5. The group labeled "None/not given" in Table 5 (TP = 0) contains treatments that were included as alternatives to systematic phonics, including language-based approaches. These treatments were either not coded by the NRP analysts, or they were used as controls. In the present study, these were coded as treatments if a separate untreated group was available as a control. In other words, our "treatments" consist of both phonics and language-based interventions.

Table 5
Breakdown of Effect Sizes by Type of Phonics Delivered in the Treatment Group*

	Outcome Set			
Level of Phonics instruction	**Average *d* (*WGT1*)**	**Sum of Weights**	**Average *d* (*WGT3*)**	**Sum of Weights**
None/not given	.356	4	.283	60
Some	.243	10	.247	127
Systematic	.514	26	.464	397
Total		40		584

*Note: Both the compromise (*WGT3*) and equal representation (*WGT1*) outcomes sets are given with sums of weights rather than *n*. However, for the *WGT1* set the sums of weights are equivalent to the number of studies. All dependent variable categories are included.

A first approximation of the efficacy of systematic phonics is thus given in Table 5 as the difference between systematic (TP = 2) and less systematic (TP = 1) phonics for which we obtained $d = .514 - .243 = .27$, using *WGT1*. This is about 30% smaller than the magnitude of the effect reported by the NRP. Table 5 also contains results for *WGT2*; however, the results are similar for both sets of outcomes. In the next section, we adjust this effect for other moderators that are correlated with the treatment variable.

Moderator Analysis

As noted above, we created some moderators and borrowed others from the NRP study Appendix G. We examined the 15 moderators below, recognizing that a single outcome could have multiple influences. For this reason, we used weighted multiple regression analysis to sort out the unique contributions of moderators in predicting effects sizes. In this analysis, we examined two sets of moderators:

Set I	**Variable Name**	**Set II**	**Variable Name**
Experimental			
Phonics	TP	Tutoring	Tutor
Language	TL	Duration	Months
Basal	TB	Standardized Test	Standard
Replacement	TR	True Experiment	Random
Control		Grade	Grade
Phonics	CP	Normal v. At Risk/LD	Normal
Language	CL	Expanded v. NRP DV categories	Tag
Basal	CB		
Replacement	CR		

Set I contains the treatment moderators based on the rubric codings of each comparison group. Set II contains other aspects of treatment including tutoring (yes or no); treatment duration; whether the instrument was standardized; whether the experiment was randomized (yes or no); grade; reader ability category; and whether the outcome fell into one of the original 7 NRP categories (yes or no).

We conducted the regression analyses with two orthogonal contrasts for degree of phonics instruction, because effect sizes may not be linear across the categories of TP as suggested in Table 5. The contrasts TP1 and TP2 were coded as:

TP1:				
	TP1 = 2/3	if	TP = 0	*No Phonics or Unknown*

	TP1 = -1/3	if	TP = 1, 2	*Some or Systematic Phonics*
TP2:				
	TP2 = 0	if	TP = 0	*No Phonics or Unknown*
	TP2 = -.5	if	TP = 1	*Some Phonics*
	TP2 = .5	if	TP = 2	*Systematic Phonics*

According to this coding, TP1 represents the difference between treatments coded as having no phonics or unknown, on the one hand, and treatments coded as having at least some phonics, on the other. The contrast TP2 represents the specific difference between treatments coded as having some phonics, and treatments having systematic phonics.

In the regression analyses below, we first entered into the equation the degree of treatment phonics (TP1 and TP2). We then entered the rest of the 14 (7 Set I + 7 Set II) variables into the regression using a forward stepwise procedure. (Note 20) We viewed this as a kind of natural competition of the variables in explaining the results, especially because we took an agnostic stance with respect to reading theory and the previous NRP results. Results for two separate regressions are reported below. In Table 6, regression coefficients are given for the *WGT1* weighting method, and in Table 7 for the *WGT3* weighting method.

Table 6
Regression Coefficients for the Analysis Weighted by *WGT1*,

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	.246	.067		3.664	.000
TP1	-.067	.110	-.036	-.610	.543
TP2	.241	.074	.186	3.262	.001
TUTOR	.399	.079	.300	5.078	.000
CL	-.320	.067	-.284	-4.795	.000
TL	.257	.083	.186	3.080	.002
STANDARD	.186	.069	.155	2.693	.008

with R^2=.322

Table 7
Regression Coefficients for the Analysis Weighted by *WGT3*,

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	.349	.050		6.998	.000
TP1	-.074	.097	-.048	-.762	.447
TP2	.188	.070	.166	2.672	.008
TUTOR	.290	.079	.231	3.665	.000
CL	-.221	.059	-.236	-3.736	.000
TL	.228	.076	.192	3.001	.003

with R^2=.199

For the *WGT1* outcome analysis, the effect of TP1 was d = -.067. This means that treatments using no phonics or an unknown degree of phonics had less of an effect than programs that did use a measurable amount of phonics. As shown in Table 6, programs using systematic phonics instruction outperformed programs using less systematic phonics with d = .241. The systematic phonics effect, however, is smaller than the effect for individual tutoring (d = .399). In addition, standardized tests tended to give larger effects (d = .186); studies in which control groups used language approaches had lower effect sizes (d = -.320); and treatments that used language approaches had larger effect sizes (d = .257). Results for *WGT3* analysis are given in Table 7. The results are similar to those in Table 6 with the effect for systematic phonics given as d = .188; the tutoring effect was moderately smaller (d = .290); and the language effects were roughly similar for CL (d = -.221) and TL (d = .228). (Note 21) Neither analysis provided evidence that randomized experiments give different results than quasi-experimental studies, or that the results differed for the NRP and the expanded set of outcomes categories.

The result for tutoring requires some discussion since it appears inconsistent with the NRP results. The unweighted effect of tutoring d = 1.09 is reported in Ehri et al. (2001, Table 2), while the effects for small group and class instruction are given as .44 and .37, respectively. Thus, the unweighted tutoring effect was documented by the NRP. When studies were weighted by size, Ehri et al. (2001, Table 1) the effect sizes for tutoring, small group, and class instruction were .57, .43, and .39, respectively. This change in NRP estimates results from the weighting scheme used, but also from the deletion of the study by Tunmer and Hoover (1993). (Note 22) In the present analysis, the deletion of this study results in a tutoring estimate of $d = .21$ ($p < .007$) while the phonics estimate is virtually unchanged.

The Tunmer and Hoover (1993) study also illustrates an important issue for interpreting the regression results. Recall that there were two treatment groups, and one untreated control group. The first treatment was the Standard Reading Recovery (SRR). It was modified by one and only one change: a systematic phonics component was added. This modified treatment was then given to the second experimental group (MRR). We coded the first group (SRR) as TP = 1 and the second (MRR) as TP = 2, recognizing the difference between the two as the best estimate of the systematic phonics effect. This is what the contrast TP2 represents. The difference between the untreated control group and the phonics groups (SRR and MRR) is the effect estimated by the first contrast TP1.

We examined residuals for the weighted regression analysis and found evidence of one outlier (standardized residual $|z| > 4.5$). This case was removed; however, this decision had very little effect on the model estimates.

Differences Between Outcome Categories

We did not explicitly examine outcomes for dependent variable categories because there were relatively few studies that contributed to any particular category. Our primary goal was to replicate results on the overall efficacy of phonics instruction. However, we did examine residuals from the weighted regression model and test for residual differences between the DV categories given in Table 1. Using an unweighted analysis (to increase n) and comparison as the unit of analysis, we found no significant differences, $F(11, 212) = .805$, $p = .635$. This implies that there were no differential effects by DV category. In particular, the average residual for Spelling was virtually zero. We do not think this result implies that phonics instruction is equally effective for all dependent variable categories, but rather that fine-grain discriminations between different types of reading outcomes require more precise data than were obtained from the phonics instruction studies.

Effects by Grade, Unit of Instruction, and Duration

We had a special interest in examining variation in effect size by grade/age. This scatter plot is given in Figure 2, in which it can be seen that in early grades systematic phonics instruction outperforms typical phonics or no/unknown phonics instruction. However, differences among these categories are small shortly after grade 3. A conservative reading of this evidence would indicate that there is no evidence that systematic phonics instruction outperforms alternative treatments after grade 3. However, the phonics indicator is confounded with other treatment variables in the early grades, and the strongest inferences about the efficacy of phonics instruction are obtained from the regression analyses. It should be kept in mind that the trends represent changes in phonics outcomes rather than changes in reading comprehension. The outcomes in Figure 2 appear to have an upward trend beginning just after grade 3. However, the existence of this trend was not verified in the regression analyses using a quadratic term for grade/age. Thus, the information in Figure 2 should be interpreted with some caution.

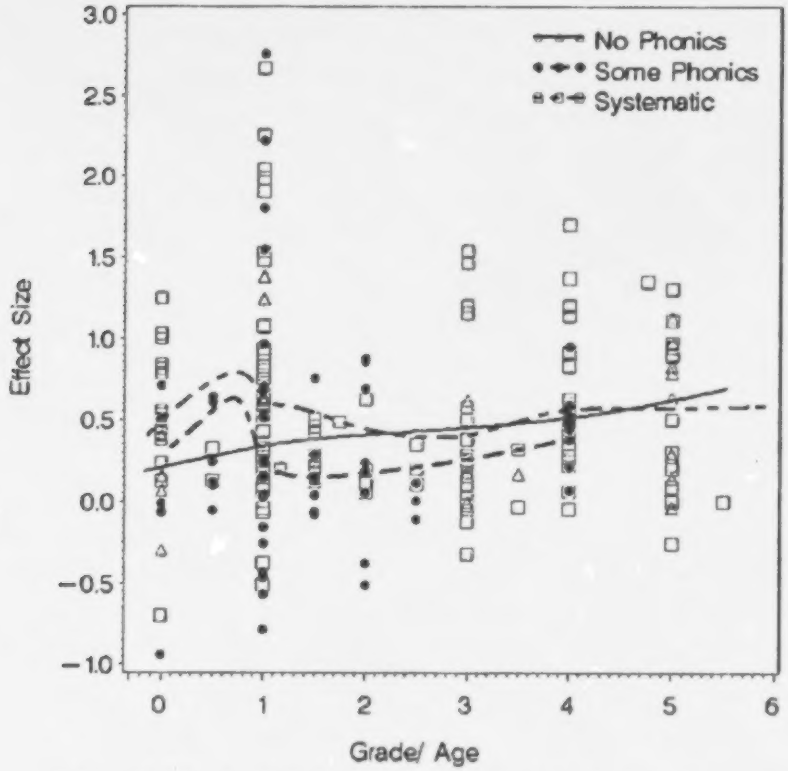


Figure 2. Effect size plotted by grade and degree of phonics instruction.
(On the horizontal axis, the point 0 (zero) represents kindergarten.)

We also plotted tutoring versus other treatment units (i.e., small group and class) in Figure 3. Here it can be seen that tutoring outperforms other instructional unit sizes across the approximate range of kindergarten to fifth grade. Furthermore, there is a suggestion, that tutoring has a greater effect in kindergarten and first grade, but also begins to increase again after third grade.

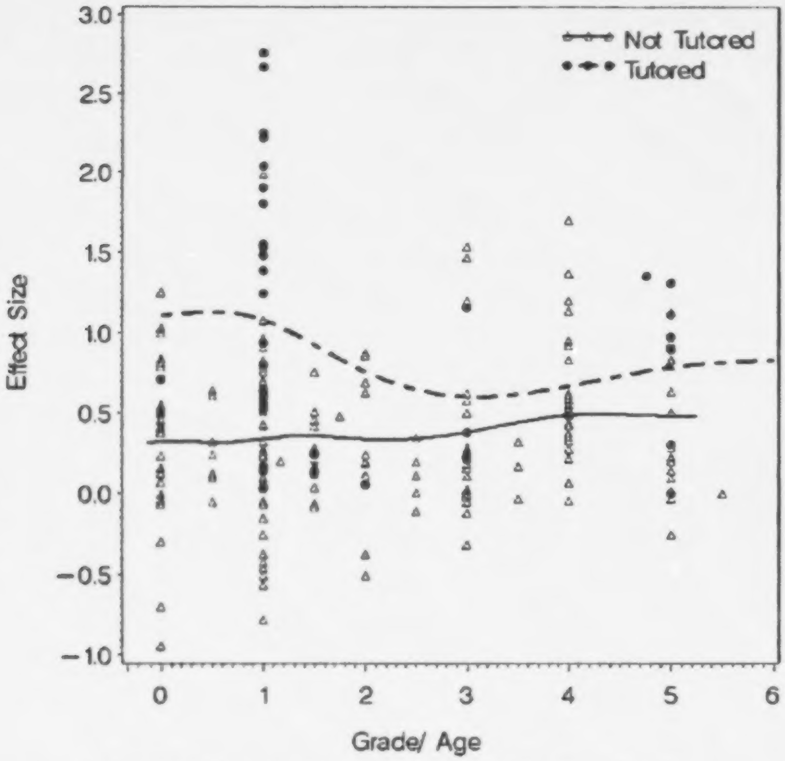


Figure 3. Effect sizes plotted by grade and unit of instruction.
(On the horizontal axis, the point 0 (zero) represents kindergarten.)

In Figure 4 effect size is plotted against the duration of treatment in months. Again the effects of tutoring are superior to those of other units of instruction, but here the effects peak at about 4 months and decline thereafter. We note that duration here denotes the chronological length of treatment and does not indicate intensity (e.g., minutes per day).

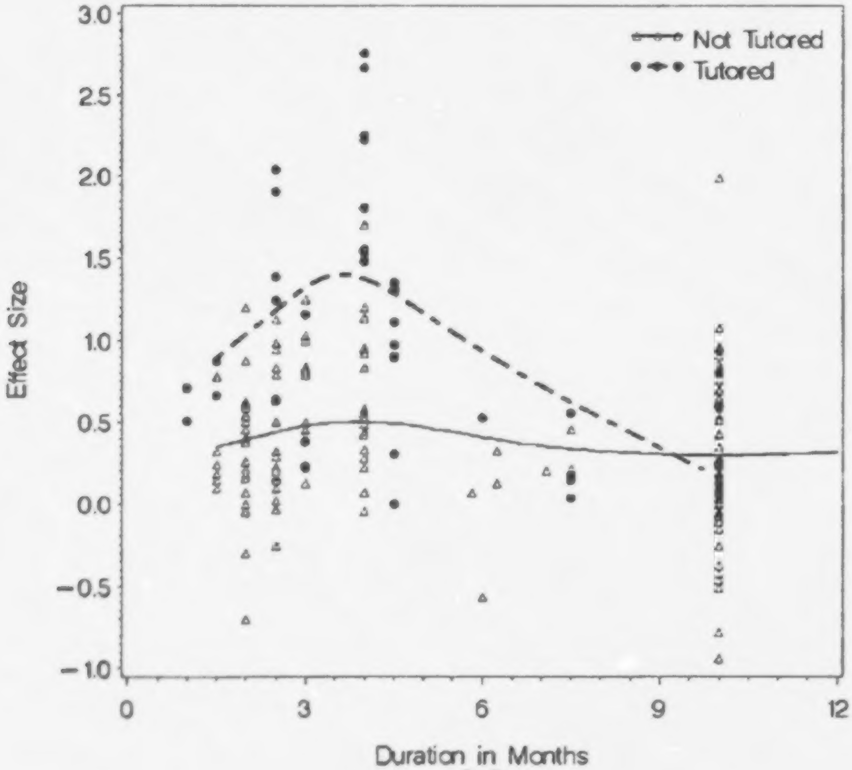


Figure 4. Effect sizes plotted by duration of treatment in months and unit of instruction.

V. Re-analysis: Discussion

Cohen (1988) is commonly cited as suggesting that an effect size of .2 is small, .5 is moderate, and .8 or above is large. However, the primary criterion for judging an effect size in educational research is its potential value for informing or benefiting educational practice. Small effect sizes can be valuable, and likewise large effect sizes can be trivial depending on the treatment and outcome in question. McCartney and Rosenthal (2000) wrote, "There are no easy conventions for determining practical importance. Just as children are best understood in context, so are effect sizes" (p. 175). Average effect sizes only provide information about whether a program works in a general sense. "A more useful question is under what circumstances do programs work best?" (McCartney and Dearing, 2002). To discover these circumstances requires that program characteristics be coded and related to effect sizes. An average effect size can also be evaluated with respect to other kinds of educational treatments. While this information does not provide a definitive rule, it does allow readers to make up their own minds about the practical significance.

Table 8
Average Effect Sizes for Instructional Methods Given by Hattie (1999)

Teaching methods	*d*	n of *d*s
Direct instruction	.82	253
Remediation/feedback	.65	146
Class environment	.56	921
Peer tutoring	.50	125
Mastery learning	.50	104
Homework	.43	110
Teacher Style	.42	*
Questioning	.41	134
Advance organisers	.37	387
Simulation & games	.34	111
Computer-assisted instruction	.31	566
Instructional media	.30	4421
Testing	.30	1817
Programmed instruction	.18	220
Audio-visual aids	.16	6060
Individualisation	.14	630
Behavioural objectives	.12	111
Team teaching	.06	41

* Not given.

In the present reanalysis, the estimated effect size for *systematic phonics* was $d = .241/.188$ (for *WGT1* and *WGT3*). This can be compared to effect sizes reported by Hattie (1999, Table 7) for various instructional methods (See Table 8). The overall average is about .4. In addition, Lipsey and Wilson (1993) examined 302 meta-analyses of a variety of psychological, educational, and behavioral interventions. Interestingly, they also found that the average treatment effect (averaging across meta-analyses) for high quality studies was .4. The largest *d*s in the present study were for tutoring (.399/.290); and use of language activities (about .288/.224). In this context, we would conclude that the advantage of systematic phonics instruction over some phonics instruction is significant, but cannot be clearly prioritized over other influences on reading skills. The regression model suggests, furthermore, that the effects of phonics, tutoring and language activities are *additive*.

It could be argued that the systematic phonics effect is actually larger than the estimate $d = .241$, and so the magnitude of the NRP estimate (about .4) is not an unreasonable expectation. However, the studies examined in this meta-analysis typically did not accurately describe the degree of phonics in the control groups. Thus, while the expectation of $d = .4$ may be plausible, it is not supported by the data. The effect size $d = -.067$ ($p > .05$) for present v. absent/unknown phonics instruction provides a cryptic message regarding alternative approaches to reading instruction. This effect is difficult to interpret because it depends on the "unknown components" of instruction. In the current study, we did not analyze this effect further. However, for teachers who currently teach some phonics, the expected benefit from a shift to systematic phonics is $d = .241/.188$. The present reanalysis suggests that tutoring and language activities are at least as effective in promoting phonics-oriented reading as systematic phonics instruction. (Note 23)

Interpretation of the Evidence on Phonics Instruction

The NRP subgroup on phonics instruction concluded that

> Findings provided solid support for the conclusion that systematic phonics instruction makes a more significant contribution to children's growth in reading than do alternative programs providing unsystematic or no phonics instruction. (NRP, 2000b, p. 2-132)

Based on our reanalysis, the evidence provides ambiguous support for this conclusion. Systematic phonics instruction did outperform treatment conditions in which a more typical or moderate level of phonics instruction was provided. But we identified tutoring and language as critical elements of a reading program in addition to phonics. The data suggest that a reading effect size has the potential to *triple* when these elements are added to systematic phonics instruction. This balance of components is critical in the early grades because the data suggest that after about third grade phonics instruction may be less effective. (Note 24) This is more-or-less consistent with the NRP finding that systematic phonics instruction is most effective in the earlier grades (NRP, 2000b, p. 2-133).

The moderator most strongly related to outcome is the unit of instruction. Tutoring showed a strong effect throughout grades 1-6 (little data are available to extrapolate further). Though shorter phonics programs tended to have larger effects, tutoring was also more effective in this instance. In programs of longer duration, the advantage of tutoring dissipated. Regarding research methodology, we found that standardized instruments (which were published and/or normed) tended to show larger effects, contrary to the expectations of the NRP analysts. This finding, however, was not consistent across the two approaches to weighting (*WGT1* and *WGT3*).

Finally, the regression results we obtained with two different approaches to weighting were roughly similar, but the deletion of one case did make a noticeable impact on the estimated effect for tutoring. This is, unfortunately, the result of a relatively small sample for conducting analyses. In this situation, there is not a single correct model for obtaining estimates, but this is not sufficient reason for ignoring the complexities of the data set. Ultimately, this problem should be resolved by examining larger samples of studies.

VI. Meta-analysis and Public Policy

In the first application of meta-analysis to research on the effectiveness of psychotherapy (Glass et al., 1981), the researchers confronted issues about research integration: how to define the population of studies to be synthesized (only published studies, only studies that met a priori standards of rigor?); how to select and measure the aspects of a study to be related to the outcomes of that study; how to classify studies and calculate their effect sizes when the primary researchers failed to report complete evidence; and how to synthesize outcomes when studies report results for varying sets of outcome measures. (Note 25) The resolutions of such questions and issues worked their way into the development of meta-analysis as a methodology that helps social scientists to distill and validate conclusions from a diverse research literature. This accumulation of research findings is not only helpful for settling disputes among researchers, but has become an important method for designing evidence-based public policies.

Meta-analysis would appear to offer great potential for objectivity and even-handedness in the synthesis of research. Prior to the 1970s, research synthesis had been fraught with bias—the reviewer selected studies that favored one perspective and cast others out, typically for *ad hoc* reasons. Because of its balanced approach, meta-analyses might resolve polarizing conflicts by making the fullest use of the research literature. The recent report from the National Reading Panel was likewise motivated in part by the desire to use the best evidence available to guide instruction in reading. Ironically, this effort has stimulated controversy regarding what constitutes evidence as well as sound research procedures.

Meta-analysis is a kind of quality control mechanism in the process of making sense of numerous individual studies. Yet criteria for the validity of a meta-analysis itself must also be considered. Is there is a general schema for producing meta-analyses that encourages the application of new knowledge? In recent years, it has become evident that a more systematic approach to meta-analysis is required in order for its original ideals to be attained. In the sections below, we explore issues of scientific due process that appear necessary for producing high quality meta-analyses, especially in areas of research laden with diverse philosophies. Included in this discussion are procedural standards, assembly of expert panels, and peer review.

Standards for Meta-Analysis

The NRP was directed to employ "rigorous research methodological standards" in carrying out its charge. However, the NRP report included a total of 7 pages (NRP, 2000a, p. 1-5 to p. 1-11) specifically addressing methodological issues (the seventh page in this section consisted of 2 references). Issues particular to phonics instruction were covered in an additional 5 pages (NRP, 2000b, p. 2-107 to p. 2-111). Altogether, less than one page is devoted to data analysis, and this contains one incorrect formula—a reference to the software used to compute the effects sizes is provided (which presumably used the correct formula). An ensuing report of the results by Ehri, Nunes, Stahl and Willows (2001) devoted just over 1 page to methodological issues beyond study selection. Perhaps this lack of attention to analytic issues was because the NRP interpreted "rigorous standards" to mean "rigorous selection criteria" for including studies, but the results of a meta-analysis depend as much on the rigor of the analytic procedures.

We think it is important for policy-oriented meta-analyses to be designed in advance with clear descriptions of basic analytic strategies. For example, the Campbell Collaborative suggests that researchers provide a rationale for why a particular effect size metric was chosen; under what conditions an effect size will be adjusted for bias; how missing data will be handled; and so forth. The Campbell Collaborative has been working on a broader set of criteria for meta-analysis that will play an increasingly important role in establishing the authoritativeness of a research synthesis. (Note 26)

Constituting Panels and Expert Review

Beyond the Campbell Collaborative principles, there would seem to be an important role of due process in selecting committees to guide meta-analyses, especially for meta-analyses that have great potential for influencing teaching practice. The Congressional bills that directed establishment of the National Reading Panel (SB 939, HR 2192) required that

> The Secretary of Education, or the Secretary's designee, and the Director of the National Institute of Child Health and Human Development, or the Director's designee, jointly shall… establish a National Panel on Early Reading Research and Effective Reading Instruction. (3:13-18)

However, the legislation itself provided only two sentences to guide selection of panel members:

> The panel shall be composed of 15 individuals, who are not officers or employees of the Federal Government. The panel shall include leading scientists in reading research, representatives of colleges of

> education, reading teachers, educational administrators, and parents. (4:4-9)

Contrast this with the selection guidelines of the Institute of Medicine (IOM), which is an institutional constituent of the National Academies of Science:

> Committees are the deliberating and authoring bodies for IOM reports, although strict institutional processes must be followed and the peer review process is independent of the committee. Most committees are consensus committees, meaning the process is designed to reach consensus on the evidence base and its implications. Where the published data are insufficient to support a conclusion, the committee may use its collective knowledge to argue for conclusions. The committee is formed by identifying the expertise and perspectives necessary to address the study topic, soliciting and receiving nominations for candidates from a wide and extensive number of sources, presenting a proposed slate and alternatives to the IOM leadership group, receiving approval from the IOM President, and formally requesting appointment from the NRC chairman. A process of seeking to identify biases and potential conflicts of interest takes place and may disqualify individuals. (Note 27)

The NICHD and Secretary of Education appear to have conducted a selection process consistent with the IOM guidelines in constituting the NRP (Note 28); however, there is no detailed description of the procedure used to choose panelists from about 300 nominees.

Visible selection procedures are important for establishing the perception of balance—that is, a diversity of theoretical and methodological perspectives—as well as actual balance. An appropriate mix of talent may facilitate a knowledge base that furthers dissemination of research findings and improves the design of new research studies. In this regard, the NRP would have benefited by formal inclusion of one or more methodologists. (Note 29) Alternatively, the research would have benefited from an officially appointed group of expert methodologists charged with translating the NRP's oversight into technically rigorous guidelines for design as well as data collection and analysis.

We could not find a description of how independent expert review of the final report was conducted. (Note 30) Moreover, a number of inconsistencies exist between the official Summary (26 pages in length) of the report and the report itself (Shanahan, 2001). If *Teaching Children to Read* had been subjected to a more scrupulous review prior to release, it would have had more potential to command a consensus. We acknowledge the severe time constraints under which the report was produced. However, the role of independent review is to verify and tighten the connections between evidence and summary conclusions. This process is intended to screen out precisely the kinds of inconsistencies and ambiguities that appear in the NRP documents.

VII. Conclusions

The impact of meta-analysis is strongly affected by two design decisions. First, the scientific due process for producing a study is critical to its acceptance. How experts are assembled and provided with resources is as important as their charge. Secondly, the science itself is important. There is no single prescription for producing meta-analyses, even though standards exist for general guidance. In spite of the expertise of research teams, time, and resources available, variability among methodological approaches is probable. Meta-analyses designed to answer controversial questions must anticipate and address this concern. One strategy might be to assemble two different teams of analysts at the onset of a study, each carrying out the five steps of meta-analysis. Another possibility may be to require methods for cross-validation in proposals in response to a formal RFP (request for proposal). Of course, such elaborate procedures are not necessary for all meta-analyses. Rather, they are most relevant to those that affect critical policy decisions, such as the studies conducted by the NRP. In any case, experts (both substantive and methodological) who do not participate in a study should provide peer review. (Note 31)

Meta-analysis is an effective method of "reading" the literature. Yet for many studies in the NRP database on phonics instruction, often little detail was given regarding treatment implementation. The NRP analysts struggled with this issue as evidenced by the number of missing study descriptors in Appendix G. Without careful description of the treatments, their implementation, and the populations of students served, it is doubtful that positive treatment effects can be understood well enough to disseminate to teachers. And without such description, it may be impossible to understand why some treatments do not work as expected. Rigorous qualitative work in reading, which the NRP is currently addressing (Manzo, 2003), has much potential to provide an effective link between theory development, program implementation, and quantitative research findings.

This reanalysis points to a number of moderator variables that may play a prominent role in designing phonics instruction. Obviously, two treatments nominally described as phonics and whole language cannot be directly compared if one uses classroom instruction while the other employs tutoring. We used regression analysis to sort out the effects of moderator variables. This provides an improvement to the one-variable breakdowns used in the NRP report. Based on the regression approach, we found that tutoring and language-based reading activities had effects at least as large as systematic phonics. In addition, the data suggest these effects are additive. These results are starkly different from the quantitative results presented in *Teaching Children to Read*, but interestingly, they are very consistent with two conclusions:

> Programs that focus too much on the teaching of letter-sounds relations and not enough on putting them to use are unlikely to be very effective. In implementing systematic phonics instruction, educators must keep the *end* [original emphasis] in mind and insure that children understand the purpose of learning letter-sounds and are able to apply their skills in their daily reading and writing activities. (NRP, 2000b, p. 2-96).

> Finally, it is important to emphasize that systematic phonics instruction should be integrated with other reading instruction to create a balanced reading program. Phonics instruction is never a total reading program. (NRP, 2000b, p. 2-97).

Despite the manifest consistency of these conclusions with the findings of the present report, the ideal role of meta-analysis—to solve controversial issues and thus to improve educational practices—was not directly fulfilled. Two independent teams of researchers arrived at substantially different interpretations of the *same* evidence.

If the NRP results are taken to mean that effective instruction in reading should focus on phonics to the exclusion of other curricular activities, instructional policies are likely to be misdirected. This interpretation of the data results from a design in which simultaneous influences on reading interventions were not adequately coded and analyzed. In particular, early literacy policies are a timely concern, especially as they are interpreted and applied in the federal Early Reading First Program. Program administrators and teachers need to understand that while "scientifically-based reading research" supports the role of phonics instruction, it also supports a strong language approach that provides individualized instruction. As federal policies are formulated around early literacy curricula and instruction, it is important not to over-emphasize one aspect of a complex process.

In our opinion, a sturdier methodology has potential to improve the estimates of the effect size in *all* substantive areas that the NRP examined. Analyses would also benefit from, indeed may require, a substantially larger sample of studies. In this effort, researchers with substantive, methodological, and classroom experience—as well as time and resources—are necessary to find studies, and to propose and test alternative design strategies. While we applaud the NRP for taking the challenging and difficult first steps in summarizing the extant knowledge on reading instruction, it is clear that more work remains to be done.

Acknowledgement

This research was completed with the generous support of The Pew Charitable Trusts. The opinions expressed in this report are those of the authors and do not necessarily reflect the views of The Pew Charitable Trusts. The authors would also like to acknowledge the contributions of Mary Lee Smith and Joanne Yatvin.

Notes

1. Results of this study were also reported in Ehri, Nunes, Stahl and Willows (2001), and Ehri, Nunes, Willows, Shuster, Yaghoub-Zadeh, and Shanahan (2001).

2. Details of this selection process are given in Section III.

3. Meta-analysis can also be performed with studies that that do not examine treatment interventions (e.g., Hunter and Schmidt, 1990). We do not consider other genres of meta-analysis herein.

4. Meta-analysis is a labor-intensive research activity. It is common to assemble research teams to facilitate the identification and coding of studies within a reasonable amount of time. However, different coders should record the same study information with a limited margin of error.

5. Readers are referred to Hunt (1997) for an accessible account of the story of meta-analysis.

6. The first estimate $d = .41$ is for outcomes at the conclusions of programs. The second estimate $d = .44$ is for end of program *or* end of school year, for programs lasting longer (Ehri et al., 2001, p. 414).

7. Fletcher and Lyon (1998) wrote "In many studies, the research was designed to evaluate the degree of explicitness required to teach word recognition skills. Instruction in word recognition skills, however, occurs along with opportunities for applications to reading and writing, exposure to literature, and other practices believed to facilitate the development of reading skills in proficient readers. This reflects one of the oldest observations of any form of teaching or training—a targeted skill cannot be learned without opportunities for practice and application." (pp. 59-60).

8. On p. 2-110 the outcome categories are given, but we could find no rationale for this particular classification.

9. Yatvin (2002) reported that "As time wound down, the effects of insufficient time and support were all too apparent. In October 1999, with a January 31 deadline looming, investigations of many of the priority topics identified by the panel a year earlier had not even begun. One of those topics was phonics, clearly the one of most interest to educational decision makers and to the public. Although the panel felt that such a study should be done, the alphabetics subcommittee, which had not quite finished its review of phonemic awareness, could not take it on at this late date. And so, contrary to the guidelines specified by NICHD at the outset, an outside researcher who had not shared in the panel's journey was commissioned to do the review" (p. 368).

10. These did not include follow up comparisons.

11. Garan (2001) shared Yatvin's concern that the NRP did not use a consistent definition of reading. Garan also criticized the NRP meta-analysis for being limited to a small number of studies and for conceptually dissimilar dependent variables. The latter two points, in our view, are problems common to both meta-analysis and narrative review. The degree to which they limit generalizability varies and cannot be determined a priori.

12. A re-examination that began at the problem formulation stage and proceeded to locating relevant studies would provide a more stringent criterion for replicability. It would also be significantly more costly. Though we skipped these two steps, we would agree that problem formulation and data collection significantly shaped the NRP's study.

13. We excluded follow up comparisons, that is, any measurements taken after post-test measurements were excluded from the analyses.

14. While some studies reported age, others reported grade. We converted all results to an approximate grade metric based on the formula grade = age – 5.

15. This formula does not appear in Cooper and Hedges (1994). See Table 16.2 on p. 237.

16. In this simple case, one divides the class-level effect size by $\sqrt{n}$.

17. We say "distinct" because each cohort involved different groups of students.

18. The data analysis described on p. 1-10 appears to use total *n*s as weights rather than the inverse variance weights described by Hedges and Olkin (1985) on pp. 86 & 110.

19. In the future "pure" statistical weights might be usefully applied when homogenous subsets of effect sizes are identified.

20. We used a highly conservative approach in the forward stepwise selection of independent variables. We required a *p*-value of .01 (PIN) to enter and a *p*-value of .05 (POUT) for removal.

21. Organized language activities were observed in about 30% of both experimental and control comparisons. Note that effective language activities in the experimental group will make the effect size *larger*, while effective language activities in the control group will make the effect size *smaller*. Thus, the two estimates logically have the opposite sign.

22. The NRP deleted one study (Tunmer and Hoover, 1993) with $d = 3.71$ in obtaining the average effect size for tutoring. The value 3.71 arose as the average of 4 effect sizes for WordID (2.94), Spelling (1.63), Nonwords (1.49), and Oral Reading (8.79). It is obvious that the last effect size is an extreme outlier, and the NRP sensibly deleted this in its computations for tutoring. We surmise that this effect size was properly deleted from other computations. We also deleted this effect size (8.71) from our computations, but we included other effect sizes from this study, which ranged from .96 to 3.18.

23. It is interesting that the effect sizes for experimental and control group language instruction are very nearly the same (taking into account reversed signs), which supports the internal design consistency of the treatment codings.

24. The gap is nearly zero at third grade, but widens somewhat at higher grades. Students in later grades do benefit, but are more likely to represent populations of reading disabled students.

25. Material on the origins of meta-analysis was provided by Mary Lee Smith in a personal communication.

26. The Campbell Collaboration is an emerging international effort that "aims to help people make well-informed decisions by preparing, maintaining, and promoting access to systematic reviews of studies on the effects of social and educational policies and practices." More information is available at http://www.campbellcollaboration.org.

27. This information is available at http://www.iom.edu/iom/iomhome.nsf/Pages/IOM+FAQs.

28. "Applicants who had taken strong stands supporting or opposing any particular approaches to reading instruction, or with a financial interest in commercial reading materials, were not considered, according to Duane Alexander, the director of the National Institute of Child Health and Human Development, who helped select the panel" (Manzo, 2000). In addition, panelists could not be employees of the Federal government.

29. Two expert consultants in methodology were introduced to the Panel in late January, 1999. It appears that both were made available to NRP members on an as needed basis. This information is available at www.nationalreadingpanel.org/NRPAbout/Panel_Meetings/01_21_99.htm. Note that the original deadline for the NRP report was January 31, 1999.

30. There appears to be a collection of documents in which the NRP's interactions are recorded. We do not know if this

archive is available for public examination (see Yatvin, 2002).

31. The Campbell group, referenced above, provides design review as a service. It does not appear to review drafts of final reports.

32. The K-3 NFT group size in the Gersten et al. (1988) study is reported as 45. Official documents give $n = 21$.

33. This model was sponsored by the Southwest Educational Development Laboratory. It stressed a developmental approach geared to children whose primary language was not English. In this approach primary language and cultural background are essential to the learning process.

34. The next five effect sizes are from Camilli (1980). They are covariance adjusted based on a modified linear model that includes a linear selection rule.

35. This model was sponsored by the City University of New York. Rather than didactic methods, direct interaction with other children was the primary method of learning. Instructional games developed skills in the areas of language, reading, and arithmetic.

36. This model was sponsored by the University of Florida. The primary emphasis was on motivating parents, and teaching them to set and attain their children's educational goals. Parents spent time as instructional assistants as well as visiting other FT parents.

37. This model was sponsored by Northeastern Illinois University. Entry language and experience of the children are built upon using a method of language elicitation focusing on the use of oral language in all curriculum areas.

38. Dissertation study, see references.

39. For all reading and spelling outcomes, the amount of growth (linear component) in each class was negatively related to initial PPVT-R standard deviations (using class as the unit of analysis).

References

Camilli, G. (1980). A Reanalysis of the Effect of Follow Through on Cognitive and Affective Development (University of Colorado, Boulder). Dissertation Abstracts International, DAI-A 41/04, p. 1366, Oct 1980.

Coles, G. (2003). *Reading the Naked Truth*. Portsmouth, NH: Heinemann.

Ehri, L., Nunes, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z., and Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36, 250-287.

Ehri. L., Nunes, S., Stahl, S., and Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*(3), 393-447.

Ehri, L. & Stahl, S. (2001). Beyond the Smoke and Mirrors: Putting Out the Fire. *Phi Delta Kappan*, 83(1), 17-20.

Fletcher, J. M., & Lyon, G. R. (1998). Reading: A research-based approach. In W. M. Evers (Ed.), *What's gone wrong in America's classrooms* (pp. 49-90). Stanford, CA: Hoover Institution Press.

Foorman, B. R., Francis, D. J., Fletcher, J.M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, 90, 1-19.

Foorman, B., Francis, D., Novy, D. & Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology*, 83(4), 456-469.

Garan, E.M. (2001). Beyond the Smoke and Mirrors. *Phi Delta Kappan*, 82(7), 500-506.

Garan, E.M. (2002). *Resisting reading mandates*. Portsmouth, NH: Heinemann.

Glass, G. V, McGaw, B., & Smith. M.L. (1981). *Meta-Analysis in Social Research*. Beverly Hills: SAGE Publications.

House, E., Glass, G., McLean, L., & Walker, D. (1978). No simple answer: Critique of the FT evaluation. *Harvard Educational Review*, 48(2), 128-160.

Hunt, M. M. *How Science Takes Stock: The Story of Meta-Analysis.* (1997). NY: Russell Sage Foundation.

Hunter, J.E. & Schmidt, F.L. (1990). *Methods of Meta-Analysis*. Newbury Park, CA: SAGE Publications.

Krashen, S. (2000, May 20). Reading Report: One Research's 'Errors and Omissions.' *Education Week*, 19(35), 48-50.

Krashen, S. (2001). More smoke and mirrors: A critique of the National Reading Panel (NRP) report on fluency. *Phi Delta Kappan*, 83(2), 118-22.

Layzer, J., & Goodson, B. (2001). *National Evaluation of Family Support Programs*. Cambridge, MA: Abt Associates, Inc.

Lipsey, M.W. & Wilson, D.B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48(12), 1181-1209.

Lovett, R., Ransby, M., Hardwick, N., Johns, M., & Donaldson, S. (1989). Can dyslexia be treated? Treatment-specific and generalized treatment effects in dyslexic children's response to remediation. *Brain and Language*, 37, 90-121.

Manzo, K.K. (1998, February 18). New National Reading Panel faulted before it's formed. *Education Week*, 27(23), 18.

Manzo, K.K. (2000, April 19). Reading Panel Urges Phonics For All in K-6. *Education Week*, 19(32), 1 & 14.

Manzo, K.K. (2002, January 30). New Panels to Form to Study Reading Research. *Education Week*, 21(20), 5.

Manzo, K.K. & Hoff, D.J. (February 5, 2003). Federal Influence Over Curriculum Exhibits Growth. *Education Week*, 22 (21), 1 & 10 & 11.

McCartney, K. & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, 71, 173-180.

McCartney, K. & Dearing, E. (2002). Evaluating Effect Sizes in the Policy Arena. *The Evaluation Exchange Newsletter*, 8 (1), 4 & 7.

National Reading Panel. (2000a). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction*. Washington, D.C.: NICHD.

National Reading Panel (2000b). Alphabetics Part II: Phonics Instruction (Chapter 2) in Report of the National Reading Panel: *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction: Reports of the Subgroups*. Rockville, MD: NICHD Clearinghouse.

Orwin, R. G. (1994). Evaluating coding decisions. Pp. 140-162 in H. Cooper & L.V. Hedges (Eds.). *The handbook of research synthesis*. New York, NY: Russel Sage.

Pressley, M. & Allington, R. (1999). Concluding reflections: What should reading research be the research of. *Issues in Education*, 5(1), 165-175.

Shadish, W.R. & C.K. Haddock (1994). Combining estimates of effect size, pp. 261-281. In Cooper, H. & L.V. Hedges (eds.), *The Handbook of Research Synthesis*, New York: Russell Sage Foundation.

Shanahan, T. (2001). Response to Elaine Garan: Teaching Should be Informed by Research, Not Authoritative Opinion. *Language Arts Journal*, 79(1), 71-72.

Tunmer, W. E., & Hoover, W. A. (1993). Phonological recording skill and beginning reading. *Reading and Writing: An Interdisciplinary Journal*, 5. 161-179.

Vickery, K.S., Reynolds, V.A., & Cochran, S.W. (1987). Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. *Annals of Dyslexia*, 37, 189-200.

Yatvin, J. (2000). Minority View. In *National Research Panel, Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction*, pp. 1-6. Washington, D.C.: NICHD.

Yatvin, J. (2002). Babes in the Woods: The Wanderings of the National Reading Panel. *Phi Delta Kappan*, *83*(5), 364-369.

About the Authors

Gregory Camilli
Rutgers University
10 Seminary Place
New Brunswick, NJ 08901

732.932.7496 X8350
camilli@rci.rutgers.edu

Gregory Camilli is Professor in the Rutgers Graduate School of Education. His interests include measurement, program evaluation, and policy issues regarding student assessment. Dr. Camilli teaches courses in statistics and psychometrics, structural equation modeling, and meta-analysis. His current research interests include school factors in mathematics achievement, technical and validity issues in high-stakes assessment, and the use of evidence in determining instructional policies.

Sadako Vargas
As Assistant Professor at Kean University, and Adjunct Professor at Touro College and Seton Hall University, Sadako Vargas has taught in the areas of research methods and occupational therapy. Her interests lie in the use of meta-analysis for investigating intervention effects in the area of rehabilitation specifically related to pediatrics and occupational therapy intervention.

Michele Yurecko
Michele Yurecko is a Ph.D. student in Educational Psychology with a concentration in educational measurement at the Graduate School of Education, Rutgers University. Her academic interests include the study of research methods and design applied to the field of education, and the intersection of educational research, testing and public policy.

Appendix A
Additional Studies

Study ID	
63	Barr, R. (1974). The effect of instruction on pupil reading strategies. *Reading Research Quarterly*, 10, 555-582. This study compared a phonics with a sight word method of instruction. Word learning tasks, word recognition, and comprehension were tested. The process by which subject were assigned to groups was not described, but it was reported that the groups did not differ in age or readiness as measured by the World Learning Tasks. Outcome variables for effect size computation were reported in terms of substitution errors on word reading tasks.
65	Peterson, M.E. & Haines, L.P. (1992). Orthographic analogy training with kindergarten children: Effects on analogy use, phonemic segmentation, and letter-sound knowledge. *Journal of Reading Behavior*, 24, 109-127. This study examined the effect of teaching orthographic analogies based on words that rhyme. Children were tested on segmentation ability, letter-sound knowledge, and reading words by analogy. Subjects were stratified on ability measures, and then assigned by odd and even numbers (sequential ranks) to treatment and control groups.
68	Gillon, G. & Dodd, B. (1997). Enhancing the phonological processing skills of children with specific reading disability. *European Journal of Disorders of Communication*, 32, 67-90. This study compared a 20-hour phonological training program to two groups tested in a previous study published in 1995. We used the original 1995 data in which a group receiving 12-hour phonological training was compared with a group receiving 12-hour semantic syntactic training. Groups were tested with the *Neale Analysis of Reading Ability – Revised*.

Appendix B
Case Studies of Three Selected Studies

Gersten, Darch, & Gleason (1988)

This study used select data from the Follow Through (FT) Planned Variation Experiment, which aimed to increase the achievement and self-concepts of children from economically disadvantaged backgrounds. To give some background, the Follow Through program was intended to pick up where Head Start ended, and maintain presumed academic gains from

Kindergarten to third grade. According to White et al. (1973, Volume II)

> [Follow Through] is intended to be a comprehensive project offering educational, medical and dental, nutritional, social, and psychological services to children previously enrolled in Head Start. Follow Through uses a strategy of "planned variation" in approaches to early elementary education, and 20 different models are being implemented in Follow Through sites across the nation. (p. 83).

Fourteen education models (i.e., different treatments) were included in the FT Evaluation (Stebbins et al., 1973), and these varied in the degree of classroom structure, basic skills, and parental involvement. One such model was Direct Instruction (DI), sponsored by the University of Oregon, College of Education. In the DI approach, behavioral methods were used with highly structured teaching materials. Teachers worked with small groups of students, and tests were frequently administered to assess children's progress.

There were two cohorts of students from East Saint Louis, Illinois. Each consisted of a treatment (FT) group receiving DI and Non-Follow Through comparison (NFT) group. One cohort was assessed from grades 1-3 ($n = 96, 45$ for FT, NFT), the other from K-3 ($n = 56, 21$ for FT, NFT). (Note 32) These were the groups providing data for the Gersten et al. (1988) study. Nationally, however, Direct Instruction was implemented at 9 other sites. Outcome measures included the Metropolitan Achievement Test with subtest scores in Word Knowledge, Spelling, Language, and Reading, among others. The NRP analysts choose to compute effect sizes for Reading ($d = .11, 28$) and Spelling ($d = -.12, .16$) for the two cohorts. The Reading effect size was classified as a measure of comprehension. The effect sizes were quite close to calculations from the present study of (.09, 27) for Reading and (-.10, .15) for Spelling. Similar national-level estimates of .14 and .12 for Reading and Spelling (for the K-3 cohort only), respectively, were given by Camilli (1980).

Overall, the results from East Saint Louis are remarkably representative of the national results, but since Direct Instruction was only 1 of 14 other models, we might ask which models showed the largest gains in Reading and Spelling. Camilli (1980) found that two models with the largest Reading effect sizes were Language Development ($d = .180$) (Note 33; Note 34) and Interdependent Learning ($d = .168$) (Note 35). In Spelling, the Parent Education (Note 36) ($d = .310$) and Cultural Linguistic (.341) (Note 37, models had the largest gains. We would add that the Direct Instruction model had the largest gain for MAT Language, Part B ($d = .327$), in which a student was required to recognize asking, telling, and incomplete sentences.

In conclusion, our statistical results are close, in this case, to those of the NRP analysts. Thus, our extended analysis of the Gersten et al. (1988) study can be taken as validation of the consistency of their methodology. However, this case study points to other aspects of the NRP study in terms of its generalizability, or external validity. It is ironic that a single study can strengthen conclusions regarding the value of phonics instruction, and yet the study was originally embedded in a larger study that provided mixed findings with regard to treatment efficacy. Though it is true that the basic skills models (Direct Instruction and Behavior Analysis) had the largest overall gains in the Follow Through experiment, the Direct Instruction model did not outperform other models for Reading or Spelling.

Data from FT models other than Direct Instruction were not included in the phonics instruction meta-analysis for several probable reasons. First, it is doubtful that reports such as those by House et al. (1978) would be identified with the NRP key word searches. It would be virtually impossible in a meta-analysis to anticipate such studies without direct knowledge of their existence. Studies like Camilli (1980) (Note 38) or the FT evaluation reports (e.g., Stebbins et al.., 1977) would not be included because they do not appear in refereed journals. However, even if such studies were located and included, a dilemma would arise because both the NFT and other FT models could serve as controls. Only if enough information were reported for comparing the level of phonics instruction in the alternative treatments could a consistent decision be made. This might be possible even though the data are about 30 years old, but such an in-depth analysis would not be economically feasible.

Tunmer, W. E., & Hoover, W. A. (1993)

This study compared the effects of three different language programs on beginning readers who had been identified as having reading difficulties. Two types of Reading Recovery programs were used for the treatment groups, and the standard intervention program was used for the control.

The first treatment group was the Standard Reading Recovery (SRR) program, which is a remedial reading program developed in New Zealand to "reduce the number of children with reading and writing difficulties." At risk children were selected and provided with 30-40 minutes per day of individual instruction by a trained teacher for a period of 12-20 weeks. Reading Recovery lessons followed the procedures developed by Clay (1985) and usually included seven activities, one of which was writing a story the child had created. Writing exercises employed phonological awareness training techniques to isolate individual sounds in familiar printed words. Incidental word analysis activities that arose from the children's responses were available after the children mastered letter identification. This instruction was given in addition to the children's regular classroom activities.

The second treatment group was the Modified Reading Recovery (MRR) program. It held the parameters of the standard program constant and then added explicit and systematic instruction in phonological recoding skills to the letter identification activities of the standard Reading Recovery program. The control group was the Standard Intervention Group. It received support services that were normally available to at risk readers, mostly funded by the (then) Chapter 1 program. Children were instructed in small groups, and instructional techniques varied greatly and included word analysis activities.

First graders with mean age of 6 years 2 months at the beginning of the school year were drawn from a pool of at risk readers from 30 schools across 13 school districts. The lowest ranked children from each school were given the Diagnostic Survey and Dolch Word Recognition tests. Three matched groups were formed from those who performed at the lowest levels on these tests. The 64 children in the two Reading Recovery treatment groups were drawn from 34 classrooms from 23 schools. The control group of 32 students was drawn from 13 classrooms in 7 schools. Classrooms were "roughly" matched on location, SES and type of classroom reading program. No significant differences were observed between the means of the three comparison groups for age and all pre-treatment measures. The study also reports that two additional control groups of 32 children each were added (p. 170), but there is no further mention of these latter groups.

For this study, the NRP analysts choose two groups, the MRR group and the Standard Intervention group. Effect sizes were then computed for 4 outcome categories: Word ID ($d = 2.94$), Spelling ($d = 1.63$), Nonwords ($d = 1.49$), and Oral Reading ($d = 8.79$). These effect sizes, especially the latter, seem very large, and this could be taken to mean that the effects of systematic phonics instruction were quite impressive. However, it should be noted that systematic phonics instruction was the key element in the MRR group that distinguished it from SRR. By comparing these two groups, we can obtain an estimate of how much improvement resulted from this modification to the standard program. We calculated these effect sizes as Word ID ($d = -.12$), Spelling ($d = -.25$), Nonwords ($d = -.12$), and Oral Reading ($d = .12$). These results indicate that these two groups performed at very similar levels.

The large SRR effect sizes may be due to either the size of the treatment unit or the RR treatment itself, but these two factors are completely confounded in this study. While the children in both Modified and Standard Reading Recovery groups received one-to-one tutoring, the children in the Standard Intervention group received small group treatment. In fact, the authors warned that

> It is important to note, however, that the highly significant results in favor of the two Reading Recovery groups over the standard intervention may not have been due to the Reading Recovery program per se (i.e., the diagnostic procedures, the format of the Reading Recovery lessons, the procedures for discontinuation) but rather to the manner in which the instruction was delivered. Reading Recovery involved one-to-one instruction, whereas the standard intervention involved instruction in small groups. (pp. 172-173)

It is arguable, in fact, that taking the authors' wisdom into account would result in an effect size for Oral Reading of $d = .12$ in contrast the NRP estimate of $d = 8.79$. Once again, we see that there is a significant issue involved in determining the definition of "control group." Whereas the NRP guidelines clearly designate the standard intervention as having the least systematic phonics instruction, it is the comparison of the MRR and SRR groups that is most germane to estimating the systematic phonics effect (in our study represented as the TP2 contrast).

Foorman, B., Francis, D., Novy, D. & Liberman, D. (1991)

This study explored the relationship among phonemic segmentation, word reading and spelling, with the intention of demonstrating the superiority of a more letter-sound (labeled "More-LS") approach of reading instruction. Children receiving less letter-sound instruction (labeled "Less-LS") were not expected to exhibit regularity effects in word reading to the same extent or at the same rate as children receiving More-LS instruction.

Two groups were selected to participate in this study. The Less-LS group was comprised of 40 students enrolled in three first grade classrooms in a Houston, Texas, public school. The More-LS group was comprised of 40 students in three first grade classrooms in two Houston parochial schools. Students in all six classes received one hour of reading instruction daily, and both groups used a basal reading series. Children enrolled in the parochial schools were younger by about 2 months on average ($p < .05$); and they had higher initial reading and PPVT (Peabody Picture Vocabulary) scores, though the latter differences were not significant. Public school classes had, on average, a PPVT standard deviation about 60% larger than that of parochial school classes.

Neither the treatment nor the control regimen was designed or manipulated by the researchers; both reflected the regular teaching habits of the individual classroom teachers. Teachers in the three public school classrooms were described as being committed to "dealing with whole words in meaningful contexts," and described themselves as using a "language experience" strategy to teach reading. The Less-LS teachers used daily story selections from the basal series *Harcourt Brace Jovanovitch Reading* to provide a theme around which instruction was based. Teachers in the three parochial school classrooms were described as being "committed to letter-sound correspondences and having children segment and blend sounds in isolation." (p. 458). Rules for relating letters and sounds, and sequenced spelling patterns were taught using *Scott, Forseman Reading, Phonics Practice Readers, Series B* and *Modern Curriculum Press Phonic Program* (a workbook). Approximately 45 of the 60 minutes devoted to reading instruction were spent on letter-sound activities. A *Scott Forseman* basal reading series was also used.

The study was approximately ten months in duration. Students were administered pre-test measures in October of first grade, with post-test measures administered the following February and May. The following tests were administered: Gates-MacGinitie Reading Test, Basic R, Form 1; Peabody Picture Vocabulary Test – Revised, Form L; a spelling test (researcher-made test consisting of 40 regular and 20 exception words), a word reading test (researcher-made test consisting of 40 regular and 20 exception words); and the 13 item Test of Auditory Analysis Skills, TAAS. There were no significant posttest group differences in TAAS mean scores or trends. There were significant differences in trends of spelling scores (both regular and exception words) and trends of word reading scores (both regular and exception words) favoring the more LS-group. In other words, the more LS-group appeared to improve at a faster rate than the Less-LS group in word reading and spelling.

For the three primary outcome variables (Word Reading, Spelling and TAAS), the researchers did not report standard deviations. In this instance, it appears that the NRP analysts used the simple standard deviation (for the effect size denominator) of class means. According to standard statistical theory, this results in an effect size that is too large by a factor of $\sqrt{n}$, where n is the number of students in the classes. The NRP effect sizes for Word ID ($d = 1.92$), Decoding ($d = 1.67$), and Spelling ($d = 2.21$) are not comparable to those of other studies in which the individual student is the basis for standard deviation calculations. In this case, we converted the effect size to the individual student metric and obtain the following: Word ID ($d = .48$), Decoding (.62), and Spelling (.49). On average, the effect sizes are 3-4 times smaller than those computed by the NRP analysts, which reflect class sizes of about 13 (for participating subjects). Moreover, approximate matching does not completely resolve the issue of what portion of the adjusted *d*s should be attributed to treatment, school type (public versus parochial), and school-by-treatment interaction.

In conclusion, the Foorman study for the most part succeeded at controlling initial differences. However, there is some evidence to suggest that the public school students lagged slightly behind their parochial school counterparts, and that individual differences in ability (PPVT-R) were somewhat larger in the public school classrooms. (Note 39)

We do not know the degree to which this initial difference may have affected posttest differences or rates of growth. However, it is clear that the effect sizes need to be adjusted to the individual student metric.

Appendix C

Description of
Vickery, K.S., Reynolds, V.A., & Cochran, S.W. (1987). Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. *Annals of Dyslexia*, 37, 189-200.

The study reports the results of a four-year study (1978 – 1981) that investigated the effect of the Multisensory Teaching Approach for Reading, Spelling and Handwriting (MTARSH) in both remedial and nonremedial classes in a public school. The study reports the result of California Achievement Test, which were administered annually in April of each year. The MTARSH was developed by adapting the individualized Orton–Gillingham-Stillman method to small homogenous groups of students. The MTARSH employs two basic decoding techniques, synthesizing phonics and memorizing whole words.

The authors report the baseline scores for each grade and the posttest scores of both remedial and nonremedial classes (separately) taken after 1,2, 3 and 4 years MTARSH instruction. The remedial classes were composed of students who qualified for Chapter 1 or special Education/LLD program, at risk of presenting reading difficulties. All other children enrolled in this school were classified as non-remedial. The MTARSH Program was employed for all students, both remedial and non-remedial, in this school ($n = 426$ during the four years covered by this study). The amount of instruction received is equal for both groups- 25-minutes per day for the first graders and 55 minutes of daily instruction for grades 2 through 6. For the remedial classes, MTARSH program was their only instruction in reading, spelling, and cursive writing. The non-remedial classes MTARSH program was taught in lieu of the regular state-adopted spelling and handwriting programs, using the supplemental reading materials and the basal readers. Although detailed instructional method and materials were different in two groups, the MTARSH method used in both classes was treated as comparable in this study.

The baseline score is from the pre-tests administered two years prior to the introduction of the MTARSH program. The intervention effect was measured by the difference between the baseline scores and the posttest scores. The analysis was conducted separately for remedial group and nonremedial groups. The NRP reports eight effect sizes for this study under general reading category. (Alphabetics, Part II. Appendix G. page: 2-174). Effect sizes are reported for 3rd 4th, 5th and 6th grades for both remedial and non-remedial groups, which yielded the 8 effect sizes computed by the NRP team. Through recalculation of the effect sizes using the formula reported (NRP Report, page 1-10) and the sample sizes reported in Appendix G, it was verified that the NRP used baseline averages as the "control group" outcome, and the one-year follow-up test averages as the "experimental" outcome. The effect sizes were reported to represent the magnitude of performance differences between the phonic instruction (Orton–Gillingham method) and regular class instruction that was provided before the MTARSH was instituted. This study examined the effect of *one* instructional method on two different populations; no control group, or other instructional method, was available for comparison. The design is clearly pre-post and does not satisfy a strict interpretation of the quasi-experimental requirement for inclusion (NRP, pp. 1-7 to 1-9).

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 1893 times since May 14, 2003

Education Policy Analysis Archives
Volume 11 Number 16 **May 14, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Networks of Schools

Robert W. McMeekin
Centro de Investigación y Desarrollo de la Educación (CIDE)
Santiago, Chile

Citation: McMeekin, R. (May 14, 2003). Networks of schools. *Education Policy Analysis Archives, 11*(16). Retrieved [date] from http://epaa.asu.edu/epaa/v11n16/.

Abstract

The study proposes: (1) that the institutional climate in schools, which includes formal rules, informal rules, mechanisms for enforcing both kinds of rules, clear objectives and an atmosphere of cooperation and trust, has a strong influence on school performance; (2) that "networks" of schools such as the Accelerated Schools Project in the U.S. and the Fe y Alegría schools in Latin America help improve school performance in a variety of ways, and have been successful in providing good education to disadvantaged children; and (3) that one of the reasons some networks are successful is that they promote the creation of sound institutional environments in member schools. The argument draws on New Institutional Economics and especially on the role of institutions inside school organizations in reducing agency problems and facilitating transactions between actors in school communities. Three examples of networks with a specific orientation toward improving equity—the Matte Schools of Santiago, Chile, the Fe y Alegría schools in multiple Latin American countries, and the Accelerated Schools Project in the U.S.—are presented and analyzed in terms of how they influence intra-organizational institutions.

Introduction

Amid all the debate about how to improve education—the controversies over vouchers, charter schools, reward-based incentives, standards and accountability—a modest but often successful approach to increasing school performance has been all but overlooked. I refer to "networks" of schools that offer distinctive approaches to providing education, and that operate in both the public and private spheres. Examples in the U.S. include Robert Slavin"s Success for All model, Theodore Sizer's Coalition of Essential Schools, the Edison Project (a private education management organization or EMO), the Accelerated Schools Project, initiated by Henry Levin and now a nationwide movement, and others. In Latin America the Fe y Alegría schools operated by the Jesuit Order in many countries of the region offer a well-known and admired example. In Chile the schools operated by the Sociedad de Instrucción Primaria (Society for Primary Instruction, SIP) provide another example on a relatively small scale.

Like some charter schools, there are networks that have been established with the explicit purpose of offering children who are "at risk" (have a high probability of failing or dropping out) an opportunity to obtain a good education. These networks appear to perform better than schools serving the same student populations but that are not members of the networks. Most networks provide a kind of yardstick or basis for comparison with regular public schools, although their main objective goes well beyond creating competitive pressures. They may be publicly financed or private, but even those that are public are to some degree outside the main bureaucracy of the education systems in which they operate. Although specifics vary, the schools in the networks must comply with certain minimum essential conditions (non-discrimination, for example) but have fairly wide latitude regarding curriculum, teaching methods, teacher recruitment and selection and school management.

Why are networks of schools more than ordinarily successful, even when they often serve disadvantaged students and when, as research has established, their costs are not significantly greater than those of other schools? Does being part of such a network confer advantages on participating schools and, if so, what are those advantages? A conceptual model, developed in the course of research in Chile and applying New Institutional Economics to school management issues, suggests some reasons. School organizations that have internal "institutions"—formal rules, informal rules and enforcement mechanisms—that are conducive to making and upholding agreements and commitments tend to perform well according to various criteria. Networks tend to promote (or protect) favorable internal institutional climates, as elaborated further below. Schools in which "the rules of the game" make it possible for teachers and other actors to interact with confidence and cooperation provide contexts in which pedagogical inputs can operate effectively. Some networks promote their own particular pedagogical approaches, which obviously have an influence on student learning. For the purposes of this study, however, pedagogical factors are left aside, as part of "other things being equal". We are concerned here with the institutional environment in the school and the way being a member of a network helps schools to develop and maintain good internal institutions. The following sections consider the nature of institutions and networks in turn.

What do we mean by institutions inside school organizations?

As Nobel laureate Douglass C. North defines the term, institutions are "the rules of the game". They include the complexes of formal and informal rules and constraints that apply in every social setting, and mechanisms for enforcing each kind of rules. They "reduce uncertainty by providing a structure to everyday life. They are a guide to human interaction" (North, 1990, p. 3-4). A good example of this way of using the term institutions is "the institution of marriage". Marriage is in some ways an economic contract, in other ways a religious sacrament, and in still other ways a set of legal rules governing property, children and matters such as insurance, taxes, rules of inheritance and so on. It is founded on and upheld by a rich complex of laws and formal rules, which can be enforced by courts, but also by long-established social norms and traditions and by the powerful influences based on the approval or disapproval of families, friends and the local culture. It is this broad-brush meaning of institutions that is used here.

A set of institutions that reduces uncertainty makes it possible for people to enter into transactions with each other. These include formal transactions such as purchase and sale of goods and services, as well as highly informal transactions—arrangements, agreements, commitments or "deals"—whereby one person undertakes to do something with a reasonably high degree of confidence that the other party will follow through on their part of the transaction. When someone contracts with another person to do something—when a principal enters into an agreement with an agent—there are costs. If there is not a well-functioning set of institutions, these agency costs can be very high. People do not have the security to enter into transactions without exercising great caution to keep from being cheated and incurring expenses of gathering information about the other party, writing and monitoring detailed contracts, paying for security or insurance and so on. If there are not sufficient safeguards, or if the costs of protecting one's interests are too high to make the transaction worthwhile, then activities that would otherwise have been useful and profitable for all parties do not get done.

The main literature on economic institutions has focused on their role at the macro level, in national economies or in major markets such as securities or commodities markets. This literature has established that institutions have a powerful effect on the costs of transactions and on economic performance. But institutions exist at an extreme micro level as well, inside organizations. There has been some study of institutions at this level. The work of Michael Jensen and William Meckling in the managerial economics literature deals in part with "the rules of the game" inside organizations (Jensen, 1998). Harvey Leibenstein's concept of "X-efficiency", especially his later work in *Beyond Economic Man* (1976) and *Inside the Firm* (1987) delves deeply into the determinants of efficiency inside organizations. Political scientist Gary Miller (1992) examines in depth the varieties of prisoner's dilemma situations that tend to arise in virtually any organizational structure or incentive scheme.

The theoretical framework used here to explore the role of networks of schools is based on the concept of institutions inside school organizations and how they influence performance. Schools constitute a large category of organizations, they tend to be organized along very similar lines and to have similar inputs and processes, yet their performance can vary greatly. There has been very little work on how institutions function inside schools (Rowan & Miskel, 1999, pp. 378-380).

When looking at institutions at this extreme micro level, it is feasible to add to the basic list of "formal rules, informal rules and enforcement mechanisms" two other factors that shape a school's institutional climate. The first is clarity of objectives, including both how clearly the objectives are articulated and how well they are understood, agreed upon and internalized by members of the school community. The second additional component of the institutional climate is the degree of trust, cooperativeness or "social capital" that obtains in the school community. These five components of an institutional climate, which are measurable, constitute a syndrome; a combination of factors that work together to influence performance. In economic terms a good set of institutions reduces principal-agent problems and the costs of transacting, thus making it feasible for beneficial transacting to take place. For a fuller explanation of this theoretical framework and the economic concepts that underlie it, as well as a description of an empirical study that measured institutions in primary schools in Chile, see McMeekin (2003).

The basic hypothesis of the theory is that, other things being equal, the more a school's institutional climate facilitates transactions between members of a school's community, the better the school's performance. "Good" institutions make it possible for teachers, students, parents and other members of the community to perform their respective tasks effectively. On the other hand, the lack of sound institutions can make it impossible for good inputs—even the best pedagogical methods and materials—to function as they should.

Three Examples of Networks of Schools

My colleagues at the Centro de Investigación y Desarrollo de la Educación (CIDE) and I became interested in networks through our study of primary schools in greater Santiago, Chile (McMeekin, Latorre. & Celedón, 2001). The private subsidized schools studied in the course of that research are all members of a 17-school network established and operated by the Sociedad de Instrucción Primaria (SIP) and best known as the "Matte Schools". Another source of interest in networks was a major study, conducted by John Swope and Marcela Latorre at CIDE, of the Fe y Alegría schools in nine Latin American countries (Swope & Latorre, 2000). Thus the study team at CIDE was alerted to the subject of networks and aware that they offered advantages to member schools in terms of fostering sound internal institutional environments. Through published sources as well as friendship and discussions over the years with Henry M. Levin, founder of the Accelerated Schools Project (ASP), I have learned enough about that project to recognize similarities between ASP and the other two networks, although superficially they may seem quite different.

One factor these networks have in common, and one reason for examining these three, is that all seek to provide a decent education to children who are disadvantaged. The Matte Schools were founded in the 19 th century to provide education to children from families with "few resources". The Fe y Alegría schools are usually established to serve children in poor rural areas "where the pavement ends". The guiding principle of the Accelerated Schools Project has been to provide a good education to students otherwise "at risk". In this sense all three networks address similar challenges and have similar objectives.

The following sections consider these three examples of networks of schools. They do not purport to offer in-depth analyses of each network. Their aim instead is to explore whether there are common factors associated with membership in these networks that help create sound institutions and contribute to good school performance.

a. The Matte Schools of the Sociedad de Instrucción Primaria of Santiago

The Sociedad de Instrucción Primaria was founded in 1856, before Chile had a public school system, with the aim of providing the children of families with "few resources" an education that would enable them to become successful, productive citizens. Its original slogan, at a time when 86 percent of Chile's population was illiterate, was "War against ignorance!" The SIP has been in continuous operation for nearly a century and a half and for much of that time the Matte family has provided strong and caring leadership as well as substantial material contributions.

There are 15 primary schools and two secondary schools in the SIP network, all in the greater Santiago area and intentionally located in low-income neighborhoods. The SIP schools, originally supported entirely from private contributions, were among the first to take advantage of the opportunity to become private, voucher-paid schools when Chile's nationwide voucher system was established in 1980. (Roughly 90 percent of enrollment in Chile is in public and private voucher-paid schools; the remaining ten percent is in elite paid private schools.)The SIP schools have always had complete autonomy to establish their own philosophy, rules, procedures, staff, and school communities and the guiding spirit of the network is very strong. These are non-sectarian schools but, since the SIP gives great emphasis to development of values and Chile is a Catholic country, a Catholic influence is clearly felt.

Private subsidized schools in Chile have higher performance than municipal schools on Chile"s national achievement tests called SIMCE (Sistema de Medición de la Calidad de la Educación). In the private subsidized subsector, the Matte schools as a group have higher scores than the average for all private subsidized schools (SIP, *Memoria, 1999*, p. 16). Table I shows comparative scores on SIMCE tests for four categories of secondary schools in 1999.

Table I
Test Scores of Tenth Grade Students, 1999

SIMCE Test	Voucher paid Municipal	Voucher paid Private	SIP 10th Grade	Elite, paid Private
Language	236	254	287	297
Mathematics	237	252	280	299

Source: Sociedad de Instrucción Primaria (SIP), *Memoria, 1999*, p. 16.

The SIP schools' performance was better than both public (municipal) and private subsidized schools and almost as good as that of the more selective and much more expensive paid private schools. Most other private subsidized schools serve upper middle class families in higher-income neighborhoods (who also self-select to attend private rather than municipal schools). The SIP schools produce higher performance on the SIMCE tests than their peer schools, even though the families they serve have lower socioeconomic status than families in other private subsidized schools. The following sections consider some of the factors that characterize the Matte Schools and contribute to their good results.

Private status. Being private clearly conveys benefits on the SIP schools, whose leaders have a clear idea of what they want to achieve. The schools in the network have freedom to establish their own "proyectos educativos" or mission statements (guided by the goals of the network itself, which are clearly stated and widely disseminated to member schools). The SIP chooses its own school directors (many of whom have "come up through the ranks" as teachers and leaders in the network) and the directors have autonomy to select and develop their own teams of teachers. These and other characteristics that private status conveys undoubtedly contribute to the success of schools in this network. Other private subsidized schools in Chile have the same opportunities conveyed by private status, but not all of them take advantage of them in the same way the SIP schools do (nor do they, in most cases, belong to networks like the SIP).

The leadership role of the SIP. "The Society", as teachers and directors refer to it, is the top authority to which the schools report. It provides the overall mission or goal statement for all the schools, while individual schools within the network establish their own mission statements on a participatory basis. This means that the principal-agent relationships between the governance authority and the schools are very close and agency problems of asymmetric information or diverging aims are minimal within the community. Another advantage the SIP provides is continuity of governance; something the municipal sector lacks.

Management support is provided by the network headquarters, which relieves member schools of performing some managerial functions such as financial management and budgeting. It also provides guidance and knowledge about management functions, interacts with the national Ministry of Education on policy matters and with the association of private schools, brings in up-to-date information on instructional and management techniques from outside the network, and provides a buffer between school-level managers and some elements of the external community that might otherwise absorb time and effort.

Team building. The directors of member schools in the SIP network are chosen by the SIP, almost always from able educators who have made their way up from classroom teacher to middle-level leadership roles to directorship. The directors then have a high degree of autonomy in building their own teams (although a new director arriving at a school inherits a staff that has been carefully chosen and trained by predecessors and the SIP system). Directors put their own personal stamp on their schools, as was evident from our interviews with five primary school directors. They dedicate substantial amounts of time to in-class observation and play active roles of professional advisor, constructive critic and strong supporter of their staff.

Staff self-select to become teachers in the SIP network. Many members of the focus groups of teachers with whom we met had been students at Matte schools or were the children of former students, or both. Clearly there is a powerful sense of mission, tradition and commitment among this staff and there is a strong feeling of belonging to a team, of both the SIP network and of the individual school. In such an environment, feelings of trust are strong.

A substantial portion of parents who send their children to SIP schools are former Matte School students or children of former students, or both. The only selection of students the schools make is to take siblings of present students or children of former students, but this in itself provides a filter that results in a community of families that understand and support the ideals and traditions of the SIP. The proportion of "legacies" varies from school to school. In some cases the neighborhoods where schools were originally sited, which were chosen because they were in low-income communities, have changed with time and shifting demographics. Some families have also moved to other communities and yet still transport their children back to the older (and often poorer) neighborhood so they can attend the school.

Parental involvement in this situation is more than ordinarily strong and parents understand and feel commitment to the schools' goals and traditions. To further strengthen the links between family and school, the SIP contracts with an independent training firm to provide "Schools for Parents" (Escuelas para Padres) that present courses on parenting skills and ways to support the school by encouraging children to do their best.

Taking all these factors into account, the community feeling in schools in the SIP network is very strong. It appears to correspond to what Bryk, Lee and Holland (1993) have observed concerning "community" in Catholic schools. In school communities such as those described here, relationships between principals and agents are positive rather than conflictive, levels of uncertainty are low, and there are low risks or costs involved in entering into transactions—between teachers and students, between peers, and other pairings—having to do with accomplishing clearly-stated goals. The environments facilitate transactions between members of the school community.

Guidance regarding rules. The SIP has its own set of rules or Reglamento Interno covering working conditions of all staff, professional advancement for teachers, directors and other senior staff, rules regarding security and risks affecting personnel and students, provisions for sanctions in the event of violation of the rules and related matters (Sociedad de Instrucción Primaria, 1998). All are expressly in accordance with Chilean labor law. These general rules are binding on schools within the network, but each individual school has its own Reglamento that deals in greater detail with rules regarding academic matters, relations between parents and the school, and responsibilities of each category of members in the school community. Members of each school community are given copies of the relevant Reglamento and parents are required to sign written statements each year attesting that they have read the rules and agreeing to abide by them and uphold their parental responsibilities.

The Society also plays a role, through written documents such as the mission statement as well as through other communication channels such as training and periodic inspections, in establishing and reinforcing the informal rules that apply throughout the network. The member schools also place heavy emphasis on establishing the norms of academic and personal behavior that teachers, students and all members of the community are expected to respect. In the informal sphere, the presence of these rules is felt as much or more through positive recognition of good performance (based not only on grades but also on being "best friend", "most improved this month", or "teacher of the month") as through negative sanctions. And in addition to explicit recognition in a variety of frequent ceremonies, teachers and directors emphasized the role played by "conversation" in providing guidance, encouragement and correction of behavior and performance where necessary.

Social capital and a climate of trust. Repeatedly we heard school directors and teachers in focus groups mention that being part of the SIP network provides a sense of security, like being part of a "family". The Society's long history and traditions contribute strongly to this, as does its emphasis on values and teamwork. Teachers inside the system may have specific disagreements with each other, with the director or some aspect of the Society's system, and directors may need to correct some aspect of a teacher's performance, but we were told that these things are worked out through "conversation", in an amicable manner, within a system of rules and procedures that all parties trust. When asked about how the system deals with teachers who are not performing well, teachers replied, "They know if they don't fit in. After a while they go away."

One of the themes among the values the SIP schools seek to teach is teamwork and cooperative behavior. And one of the explicitly recognized means of teaching values is through modeling, in which school directors and teachers engage actively. The formal and informal rules serve to establish an environment where all members feel secure, agency problems are minimal and transaction costs are low. While there is a strong sense of discipline, this is maintained through moral pressure and carefully managed so that individuals are not shamed or hurt.

Because the directors and teachers tend to progress through their careers within the system (and often are former students), they are strongly socialized into the SIP system of values. There is agreement and consistency in the educational communities. Having the Society as an umbrella organization shelters the school from shifting influences from external sources.

Material benefits. The SIP provides network management and leadership functions that relieve its schools of some administrative burdens. There are also services, especially teacher training and evaluation and activities such as "schools for parents", that are provided centrally. The Society's staff is well informed about the findings of educational research and best practice throughout Latin America and the world, and make this information available to the individual schools, which would otherwise not have access to it. It is a source of guidance and wise advice for its member schools.

By far the largest share of income comes from the government via various categories of transfer payments, the largest being the subventions or vouchers. The second largest source of income is payments by parents through the system of shared financing or "financiamiento compartido" that operates in 11 of the 17 schools. Contributions from the Society itself (members of the Board and senior staff are required to make contributions) and donations from third parties help fill the gap between income and expenses.

The Society manages its funds centrally and allocates them between schools in the network. It has also been instrumental in arranging for shared financing in those schools where parents have agreed to this measure, and in raising funds from private donors.

b. The Fe y Alegría schools of Latin America

The Fe y Alegria (FYA) network was founded by Jesuit Father José María Vélaz in Venezuela in 1955. It now operates in fourteen countries of Latin America plus Spain.

The Latin American countries in which FYA operates are: Argentina, Bolivia, Brazil, Colombia, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Paraguay, Panama, Peru, Spain and Venezuela. Some country programs are quite small. The system was created to provide education to underprivileged children. Most of the FYA schools are located in rural areas but some are in or near urban slums. In each country a National Office coordinates the schools in its network and these national federations enjoy a high degree of autonomy. Over all coordination is provided by a headquarters in Venezuela. The statement of guiding principles to which all schools in the system adhere is the International Mission Statement of FYA. Each National Office develops its own mission statement, and each school within a federation has its own, locally developed mission statement. This, then, is a network of networks.

The schools' main aim is to reduce the high levels of repetition and dropouts that plague education in most Latin American countries. They strive to provide a good education and to assure that students complete at least the basic cycle of schooling. Student retention in school is the main indicator of performance, since most countries of the region lack standardized tests that would make it possible to evaluate schools on the basis of test scores. The FYA network is notably successful in reducing repetition and dropouts. The CIDE study of FYA in nine countries showed that in a majority of countries, FYA schools serving poor children achieved better results than the national averages for all public schools in terms of "opportune progress" (meaning completing the primary cycle within the expected number of years), lower repetition rates, lower definitive dropout rates and higher overall student retention rates (Swope & Latorre, 2000, pp. 102-114). Retention rates for girls were higher than for boys in FYA schools. Table 2 shows gross repetition rates and cohort dropout rates for FYA schools and averages for other public schools in nine Latin American countries.

Table 2
Comparison of Repetition and Dropout Rates in
FYA Schools and other Public Schools in
Nine Countries of Latin America

	Gross Repetition Rates*		Gross Dropout Rates**	
Country	**FYA**	**Public Schools**	**FYA**	**Public Schools**
Peru	25.43	32.48	9.86	25.72
Bolivia	20.35	72.88	9.00	26.80
Venezuela	22.03	40.24	16.18	38.65
Nicaragua	24.07	39.11	13.75	10.00
Ecuador	7.21	12.80	29.20	38.60
Guatemala	20.49	18.04	22.30	38.20
Colombia	21.33	19.21	10.50	8.00
El Salvador	29.03	20.20	39.70	40.40
Paraguay	27.44	33.88	8.37	5.00

* Traditional gross grade repetition rates, equal to the accrued number of repeating students registered by a cohort as a percentage of enrollment in the cohort's beginning year.
** Includes dropouts from all causes including, for example, moving away.
Source: Swope & Latorre (2000), pp. 104-105.

Recall that the FYA schools, intentionally located in poor rural areas, are compared here with the average for all primary schools in their countries.

FYA schools are both public and private. The proportions vary from country to country but a majority of FYA schools are public in every country. In most countries the schools enjoy a high degree of autonomy, even though they operate with public funding and under public control. This autonomy is not always fully respected but when it is, it enables the National Offices to choose their own school directors and teams of teachers, without interference from the state or teachers' unions, and to keep teachers from being transferred out of the schools unless they choose to go (Codina, 1994, p. 333). Autonomy gives FYA schools a chance to build their own staff, which leads to like-minded teams, committed to the same goals. This in turn tends to reduce agency problems and the transaction costs involved in making agreements within the team, including between directors and staff and between peers.

The CIDE study found that the FYA schools do not differ substantially from public schools and that they are "a more efficient type of public school" (Swope & Latorre, 2000, p. 159). The main source of funding for FYA schools is government transfers to cover teacher salaries, which is by far the largest expenditure item for basic education. But in order to provide additional expenditures for textbooks and teaching materials, teacher training and (whenever possible) for educational innovations, it is necessary to raise funds from other sources. This is done through development of small school improvement projects, which become focal points for community involvement as well as channels for communication between the individual schools and higher levels in the FYA organization. The international and national offices of FYA help schools prepare projects and present them to donors for financing. Projects are thus both a source of additional financing and a means of forging stronger links between the school, its community and the network.

What are the reasons FYA schools achieve better student retention than their essentially similar public counterparts? The affiliation with the Catholic Church and the mystique and charisma that this provides is undoubtedly a factor, as is the high degree of autonomy mentioned above. Pedagogical inputs are also important (although the network encourages schools to develop their own models and philosophies of education in consultation with parents and the community). It is clear, however, that the network provides an overall vision or set of objectives as well as management practices that promote the development of a desirable institutional climate in its member schools. Such an institutional environment makes it easier for members of school communities to enter into the kinds of informal transactions that contribute to effective teaching and studying and good performance.

Mission and objectives. The international mission of FYA plays a powerful role in shaping the schools in the network. The mission statement helps explain the network's aims and convince communities to invite FYA to establish a school in their neighborhood. As Swope and Latorre (2000) emphasize, the language of the mission statement is "value laden" and "aimed at spreading an almost missionary message" (p.26). Examples of emotive language include:

- Popular education movement
- Commitment
- Construction of a more fraternal and fair society
- Educational action and social advancement
- Option for the poor
- Reach the poorest sectors of our cities
- Quest for a more equitable world
- A child without a school is everyone's problem
- Fe y Alegria starts where the pavement ends (p. 26).

Such evocative language, plus a single clear focus on keeping children in school, articulates the network's objectives and helps principals and agents understand each other, communicate effectively and bring forth the best efforts of all members of the network and its individual school communities.

Creation of a school community. Local populations must invite FYA to establish a school in their neighborhood, which provides the starting point for close relationships with parents and community leadership. FYA chooses its school directors, who then have the latitude to choose their own cadre of teachers. FYA helps provide teacher training in its aims, techniques and methods. It also manages to reduce teacher turnover, which is a chronic problem in schools in Latin America, especially schools in rural and poverty areas. Low teacher turnover further contributes to a sense of bonding and trust. (It is noteworthy that FYA does not attempt to provide economic incentives for teachers, yet teachers' commitment is very high.) These factors create feelings of community and teamwork.

Involvement of parents and the external community. Once a community elects to invite FYA to establish a school, it becomes closely involved with the school. One of the key "student-retention strategies" FYA uses is programs to involve the community and families. Activities promoting involvement include "workshops… for parents as a way of having them gain a better understanding of the problems that their children face", "increasing awareness of the importance of education", "strategies to get parents to support school work in the home", "programmed home visits", and "community involvement strategies, especially those designed to involve parents in the education of their children" (Swope & Latorre, 2000, pp. 120-121).

Formal and informal rules. Because there are obvious differences between the situations of FYA schools in different countries, as well as great diversity among schools in any one country, the FYA network does not present guidelines for formal rules in school. In general these schools, affiliated with a Catholic organization, tend to be orderly and to emphasize values and cooperation. The most important role of FYA is in the area of informal rules and enforcement mechanisms, especially in the relationships between school directors and teachers and between teachers as peers.

Environment of cooperation and trust. Schools in the FYA network are managed privately, although they receive public subsidy, and this facilitates establishing school communities in which teachers remain for extended periods, professional relationships between directors and teachers tend to be close and positive, parents are brought into contact with the school and encouraged to support it and motivate their children to study well and remain in school. FYA's strong sense of mission contributes to feelings of solidarity and participation in a worthy undertaking. Writing of the "key elements that determine the quality and efficiency of FYA", Father Gabriel Codina, S. J., head of FYA in Bolivia, says "The most important thing, without question, is the sense of 'mission', the charisma of FYA itself and its choice and spirit of service to the poor. This provides a 'value added' that is a particular characteristic of FYA in all of Latin America" (Codina, 1994, p. 344).

c. The Accelerated Schools Project

One of the best known and most successful networks of schools in the U.S. is the Accelerated Schools Project (ASP), initiated by Henry M. Levin and colleagues at Stanford University in two pilot schools and now encompassing over 1,000 schools in 41 U.S. states. Schools in the ASP serve student populations that were "at risk" and were performing poorly before being incorporated into the network. Sixth grade students in the schools were reading one to two years below grade level when the schools opted to become Accelerated Schools. ASP's basic concept is that, in such situations, children need schooling that is more intensive (accelerated) than other schools rather than "dumbed down" because the students are below grade level in reading and math. The essential aim is that, by the time students complete primary or middle school, they should be reading and generally performing as well as (if not better than) the average for the district. The ASP applies techniques borrowed from schools for gifted and talented students, with the idea of providing for students at risk an education that anyone would want for their own children.

In most cases schools in this network are regular public schools and not special or charter schools. The ASP obtains permission and support from the local public education authority to put its philosophy and techniques into practice (and for this to happen, the local authority must support the idea). The schools are allowed to build their own teams of personnel and to practice their own pedagogical approaches with minimum intervention from above.

Building on the experience in the "pioneer schools" that were the first members of the network, the ASP has developed a philosophy based on three principles, a set of nine values, and instructional methods called "powerful learning" that integrate curriculum, instructional techniques and organization (Finnan, St. John, McCarthy & Slovacek, 1996, pp. 15-19, 297-301). The principles are "(a) unity of purpose, (b) empowerment with responsibility, and (c) building on strengths" (Finnan *et al.*, p. 15). The values include "the school as a center of expertise, equity, community, risk taking, experimentation, reflection, participation, trust, and communication" (Finnan *et al.*, pp. 17-18). A powerful learning situation "is one that incorporates changes in school organization, climate, curriculum, and instructional strategies to build on the strengths of students, staff and community to create optimal learning results" (Finnan *et al.*, p. 18). This brief description cannot do justice to the richness of the principles, values and learning approach that have evolved since the project began in 1989. It does, however, suggest the vision and guidance the Accelerated Schools network provides to its members, and the way the network fosters strong institutions within its member schools. Detailed information on how these overarching ideas are put into effect is found in Hopfenberg, Levin, Brunner, Chase, Christensen, Keller, Moore, Rodríguez and Soler (1993).

Becoming an Accelerated School. One of the outstanding features of the ASP is the heavy investment—especially of time, expertise of ASP staff and coaches, and hard work of people at the level of the school that wants to join the network—that precedes full membership in the project. Before schools can join the network there is what Finnan *et al.* (1996, pp. 82-103) call a "courtship phase", involving extensive introspection and the development of strong commitment on the part of all members of the school community—including the local board, school leadership and staff and parents—to the philosophy. This is one means by which the ASP achieves a high degree of agreement between principals and agents.

Once the decision is made to join the project, a period of training and preparation begins. School district officials, school staff, parents and community members participate in preparing for the school to become an Accelerated School. Teachers have considerable voice and latitude to put their ideas into effect, which builds their sense of participation and their commitment to the program that emerges. The processes involved in implementing the project's concept of "powerful learning" create an environment in the classroom that is conducive to learning. Finnan and her colleagues found that teachers in ASP schools have a feeling "that collaborative practices that come about as a result of the Accelerated Schools model have led to their individual professional growth and increased communication and sharing among colleagues. A by-product of this collaboration and sharing of philosophy has been a greater feeling of support from peers, camaraderie, and accountability among the staff to provide a better education for all of the students" (p. 283).

Evaluation of ASP schools. Accelerated schools have generally achieved their aims of bringing students up to grade level in reading and mathematics and to levels of performance as good as or better than other schools in their district or area. Since their target populations are defined by being at risk, and would tend to have below-standard performance without the intervention of the ASP, evaluation on the basis of test scores must be done with care, since cross-section comparison is not a valid way of assessing the value added they provide. Scores on standardized tests do not capture the full range of the effects of ASP membership. These effects include not only gains in student achievement but also improvements in attendance rates, reductions in repetition, student suspensions and vandalism; increased parental involvement and numbers of ASP students who meet the criteria for traditional gifted and talented programs. A report by the New American Schools Corporation (NAS) summarizes findings on the effectiveness of the ASP (New American Schools, n. d.) Examples included fewer disciplinary referrals to the school principal, increased student attendance, reduction in the number of suspensions, and decreases in numbers of students retained in their grade. In the area of family and community involvement, NAS reports that a number of Accelerated Schools experienced large increases in parental involvement, including participation in decision-making and volunteer activity, as well as more student participation in community and extracurricular activities.

There are indications that ASP schools produce good results in terms of test results. The National Center for the Accelerated Schools Project publishes information on the "Accomplishment of Accelerated Schools" on its website, showing that ASP schools in Ohio, Tennessee and Texas produced higher reading and math scores than the average for schools in the same districts where the ASP schools are located (National Center for the Accelerated Schools Project, 2000).

An independent evaluation of ASP was conducted, with funding from The Ford Foundation, by the Manpower Demonstration Research Corporation (Bloom, Ham, Melton, & O'Brien, 2001). The MDRC study used an innovative methodology that compared the performance of schools in a sample during five years of implementation of the ASP model with base-line scores on reading and math tests during the three years prior to implementation and projections of what each school's average test scores would have been without ASP. The study found that, after five years of implementation, the ASP schools in the sample had higher third-grade reading and math scores than a set of base-line scores by a statistically significant amount.

On average, the schools increased their students' relative performance by 6 percentile points in reading and 7 percentile points in math during the five-year follow-up period…. The lowest-performing schools at baseline experienced the largest improvements in test scores" (p. 62). "Implementation of the Accelerated Schools reform was a difficult, time-consuming process. But schools that stuck with the reform were able to improve the school environment appreciably, especially with regard to organizational culture and decision-making. These environmental improvements were followed by increases in students' test scores, although it was not until the fifth year after the initiative was launched that average scores rose above their baseline levels by a statistically significant amount" (p. 74).

The researchers are cautious not to draw too strong conclusions from their findings and spell out a number of caveats concerning the study and its interpretation. One must bear in mind, however, that students in these schools had been performing much below grade level prior to implementation of ASP, and that the program succeeded in bringing their average scores significantly closer to regional or national averages.

The ASP and institutions in schools. The first of the three principles of Accelerated Schools—unity of purpose—is closely related to clarity of goals or the objective function, with all that implies for reducing principal-agent problems. The ASP's values of equity, community, risk taking, experimentation, reflection, participation, trust, and communication reverberate strongly with the characteristics of schools with positive institutional environments. Requiring strong initial commitment on the part of the local authority, school director and teachers, and parents before a school becomes a member of the ASP is consistent with clarity of goals and also has overtones of contracting.

Clearly the emphasis on community has an effect on the institutional environment of ASP schools (reminding one of the findings of Bryk, Lee and Holland about community in Catholic schools). The ASP clarifies goals, builds powerful community feeling and commitment, and promotes an atmosphere of cooperation and trust. It has less explicit involvement in laying down formal rules than the Matte schools (in part because it leaves this to each school to establish for itself) but the development of a sense of community necessarily contributes to formation of informal rules and enforcement mechanisms. This is a different kind of network; one that transmits a powerful ideology to its member schools and provides information, training and oversight to help them implement an effective program. The ASP philosophy creates school environments in which relationships between principals and agents are based on shared goals and trust. This tends to reduce uncertainty in interactions or "transactions" between community members and makes it easier (less "costly" in terms of risk and uncertainty) for members to enter into arrangements having to do with how much energy and effort they will expend in pursuit of their school's goals.

Conclusions about networks

The three examples discussed above were chosen because of their similarities to each other, especially in their focus on equity objectives. Not all networks possess all the characteristics observed here, nor do all tend to promote positive institutional environments in the same way. On the basis of these examples, however, and information on other networks, it is possible to suggest the following ways in which networks with similar, equity-oriented objectives operate to promote positive institutional climates within their member schools. Networks tend to have the following characteristics:

- *Short, clear principal-agent links or relationships.* Network operators (who are often private groups) maintain close relationships with the schools and provide their vision, guidance and supervision directly to the schools. There tends to be a high degree of consensus between the network's leadership (or operating authority) and the schools in the network.
- *A strong sense of mission.* Networks are usually formed because some leader or group seeks to accomplish an educational mission. Those mentioned in the sections above have all sought to improve education for children of families in or near poverty. Such a mission attracts participants—whether parents, teachers or others—who agree with the mission and are committed to working toward its accomplishment.
- *Clarity of objectives.* A general characteristic of networks of schools is a set of clear objectives, which are communicated effectively to all members the educational communities within the network, including parents. Clear objectives (that minimize problems in principal-agent relationships) are characteristic of schools with favorable institutional climates.
- *Social capital.* The three examples of networks described above all provide a sense of belonging to a parent organization within which there is trust, cooperation and mutual support, and this tends to be transmitted down to the individual schools as well.
- *Parent and community involvement.* All the networks discussed give high priority to involving parents and, to varying degrees, the surrounding community in school affairs. Parents are encouraged to support the school's objectives and motivate their children to study hard and make genuine effort.
- *Consensus within the community.* In the process of establishing a new school, or bringing an existing school into a network, some networks require that all parties to the school (including the public authority within which the school operates) understand and "buy into" the objectives and procedures of the network. The Accelerated School Project follows a procedure whereby it consults with members of a school's community over a period of up to two years, during which it establishes firm commitments, before a school becomes a member.
- *Contribution to establishing rules.* Networks help establish a clear set of formal rules, and may lay down guidelines for their member schools to follow, but individual member schools usually have a high degree of autonomy to clarify their own rules. In small-scale networks such as the Matte Schools, informal rules seem to be communicated within the network, especially those regarding the roles and responsibilities of teachers, parents and school directors. All the Matte Schools have a "reglamento interno" or set of written rules, and parents in this network are required to sign statements specifying that they have received the formal rules, understand them and agree to abide by them.
- *Continuity.* Networks provide a degree of continuity regarding objectives, leadership, rules and pedagogical orientation, and constitute a buffer between the school and political authorities, or a stabilizing force in times of political change or shifting educational fashions. They assume the role of the Board or governing authority, and the institutional relationship between the network/Board and the school is thus consistent and positive. Principal-agent relations between external authority and school are clear, strong and consistent.
- *Guidance and supervision.* One of the benefits networks provide is sharing experience within the network and making information (such as up-to-date knowledge on relevant research findings or proven practice improvements) available to network members. In addition to the value of the content itself, such communication tends to provide professional stimulation. The network leadership constitutes a respected supervisory authority.
- *Administrative functions.* Some networks assume responsibility for certain administrative activities, including relationships with the next higher administrative authority, screening and hiring teachers into a pool, central purchasing, or negotiations with unions, if this is relevant. This relieves schools of some administrative burdens.
- *Evaluation of performance.* Networks may establish their own system of evaluating member schools and their staff according to their own criteria and standards. The Matte Schools network has a carefully-designed system that teachers understand and appear to trust.
- *Rewards and recognition.* Networks may establish systems of awards for good performance to individual member schools or teachers. (Since winning teachers are chosen from throughout the network rather than from within an individual school, this form of reward does not create undesirable competition, distortion of behavior or resentment toward either the winners or those who choose them.) The rewards may be monetary but many are based only on recognition, which provides an incentive many teachers value highly.
- *Inputs of resources.* Most networks provide resources in the form of leadership and guidance. Some also offer curricular guidance, materials, training and, in some instances, material resources in money or in kind. In most cases the financial resources available are limited, but some networks are adept at raising additional funding.

The characteristics of the networks examined here have strong parallels with: (a) the findings of several bodies of research on what makes a good school (including research on Effective Schools, Catholic schools, school culture, school leadership, site-based management and parental involvement), and (b) the characteristics of schools with good institutional environments.

Other authors have recognized the way networks can contribute to improving performance of their member schools. In a section titled "The Importance of Networks", Darling-Hammond, Ancess and Falk (1995) discuss the way in a local network in New York influences the five schools practicing authentic assessment they describe in their book. All five are members of the Coalition of Essential Schools (CES) network and four of the five are also members of the Center for Collaborative Education (CCE), the New York City affiliate of CES. The authors make the following comments about networks:

> While these networks have provided common ground for sharing practice and for exploring new possibilities, each school has interpreted and enacted the CES principles in quite different, contextually appropriate ways… [This] underscores the importance of ensuring that practitioners invent models – rather than replicate models—that are embedded in and embody their knowledge of their local contexts…. As a network enables practitioners to consider these issues across schooling levels, each learns important strategies from the other. Elementary school work is strengthened as communities reflect on their values and purposes, articulating their expectations for what students should be able to do and developing public criteria for their standards and expectations…. The network is a vehicle for reconsideration of practice from many vantage points, centering around a hub of common values supporting learner-centered practice. This makes connections between like-minded practitioners both possible and mutually profitable…. it expands the possibilities for the kinds of conversations that practitioners need to have if teaching, assessment and school structure are to be organized for school success…. (Darling-Hammond, Ancess & Falk, 1995, pp. 266-268.)

Linda Nathan and Larry Myatt (1998) find that, on the basis of their experience managing a "pilot school" within a Massachusetts school district, being part of a local network of such pilot schools had benefits for school leaders.

> Networks are powerful tools for schools because they provide meaningful feedback as well as greater and better accountability than bureaucracies. Friendly, finely-tuned feedback from a number of sources is indispensable to a good school…. Networks of schools and educators provide the opportunity to grow and learn with and from others who share a clear purpose and whose work we know, trust and respect. Participation in our pilot network is loose and flexible, but with appropriate degrees of critical friendship at a negotiated pace and style. Feedback from peer schools can have a tough edge to it and is perhaps taken more to heart because it comes from those who know young people and the profession firsthand… Membership in a national network, such as the Coalition of Essential Schools, gives [our school] access not only to an invaluable template of school reorganization but also to the voice of the practitioner—with a democratic and multicultural tone. (Nathan & Myatt, 1998, pp. 283-284.)

The subject of networks is attracting increasing interest. There are various kinds of networks, including those that provide a set of ideas or principles (such as the Coalition of Essential Schools and others), those that require schools to adopt a specific instructional approach or model (such as Success for All), profit-making networks or EMOs such as the Edison Project and others.

The statements of both Darling-Hammond, Ancess and Falk (1995) and Nathan and Myatt (1998) mention small local networks of schools that have something in common (e.g. pilot schools) that provide mutual support (including moral and political support) and play an important role in facilitating the exchange of ideas and information between like-minded professionals. In both cases the local networks were part of the larger Coalition of Essential Schools network.

A RAND Corporation study of the New American Schools (NAS) network analyzed the performance of seven models of "whole-school reform" that had many of the characteristics of networks (Berends, Bodilly & Kirby, 2002). NAS itself is a network of networks that provides information and a menu of alternative models of restructuring—as well as detailed information on model design, some funding and other inputs—to schools that are seeking to reform or restructure. Once a school has decided on the model it wants to follow, NAS provides additional information, advice and other inputs that help the school implement it.

The models of whole-school reform that RAND studied were: Purpose Centered Education of Audrey Cohen College; Authentic Teaching, Learning and Asessment for All Schools; Co-NECT Schools; Expeditionary Learning Outward Bound; Modern Red Schoolhouse; National Alliance for Restructuring Education, and Roots and Wings (Berends, Bodily & Kirby, 2002, pp. 37-41). These models had some of the characteristics of networks. They were "external change agents" that helped schools implement the models, but they did not offer the same articulation of a mission or set of objectives, nor the same sense of being part of a broader community of like-minded professionals to which Nathan and Myatt (1998) refer. Nor did they have the same kind of impact on institutions within the schools that networks such as the Matte Schools, Fe y Alegría or Accelerated Schools have.

Different networks have different objectives and operate in different ways. Lieberman and Grolnick (1966) provide a review of different types of networks, focusing on the processes involved in starting and operating networks. Their study raises a number of penetrating questions about networks and their role in creating communities in which professionals with common interests can share experiences and provide valuable information and mutual support. It illustrates the diversity of networks but does not analyze in depth the ways networks influence their member schools.

Not all networks have all the characteristics of the three equity-oriented networks this paper considers, nor do they have the same influence on institutional environments within their member schools. Research on networks in general has reached guardedly positive conclusions about their ability to improve school performance. The three networks examined in this paper appear to help their member schools to perform as well as or better than other schools in their jurisdictions and to provide disadvantaged children with a better education than they would otherwise obtain. This paper argues that one way they do this is through their influence on institutions within schools.

The conceptual framework of institutions and their influence on school performance, described briefly in Section B above, helps explain the benefits of network membership. These three networks provide examples of how promoting strong institutions within schools can improve the education of disadvantaged children. Both subjects—networks and institutions in schools—deserve further study.

Acknowledgement

The author wishes to thank the Spencer Foundation for support for developing the model presented here and research on the Matte Schools network, Patricia Matte and personnel of the Sociedad de Instrucción Primaria, John Swope, Marcela Latorre and two anonymous referees.

References

Berends, M., Bodilly, S. J., & Kirby, S. N. (2002). *Facing the new challenges of whole-school reform: New American Schools after a decade*. Washington: RAND Corporation.

Bloom, H. S., Ham, S., Melton, L. & O'Brien, J. (2001) *Evaluating the Accelerated Schools approach: A look at early implementation and impacts on student achievement in eight elementary schools*. New York and Oakland, CA: Manpower Demonstration Research Corporation. Retrieved from: www.mdrc.org/reports2001/AcceleratedSchools/accSchools-overview.htm

Bryk, A., Lee, V., & Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.

Codina, G. (S.J.) (1994). 'La experiencia de Fe y Alegria', in Gajardo, M. (ed), *Cooperación internacional y desarrollo de la educación*. Santiago: Agencia de Cooperación Internacional de Chile.

Darling-Hammond, L., Ancess, L. J. & Falk, B. (1995). *Authentic assessment in action*. New York: Teachers College Press.

Finnan, C., St. John, E. P., McCarthy, J., & Slovacek, S. P. (1996). *Accelerated Schools in action: Lessons from the field*. Thousand Oaks, CA: Corwin Press.

Hopfenberg, W. S., Levin, H. M., Brunner, I., Chase, C., Christensen, S. G., Keller, B, M., Moore, M., Rodriguez, G. & Soler, P. (1993). *The Accelerated Schools resource guide*. San Francisco: Jossey-Bass.

Jensen, M. C. (1998). *Foundations of organizational strategy*. Cambridge, MA: Harvard University Press.

Leibenstein, H. (1976). *Beyond economic man*. Cambridge, MA: Harvard University Press.

Leibenstein, H. (1987). *Inside the firm*. Cambridge, MA: Harvard University Press.

Lieberman, A. & Grolnick, M. (1996). Networks and reform in American education. *Teachers College Record 98* (1), 7-45.

McMeekin, R. (2003). *Incentives to improve education: A new perspective*. Cheltenham, UK & Northampton, MA: Edward Elgar Publishing.

McMeekin, R., Latorre. M. & Celedón, F. (2001). Institutions within School Organizations: Looking Inside the Black Box. Portland, Oregon: ERIC Clearinghouse on Educational Management, ERIC Document Reproduction Service No. ED453571.

Miller, G. (1992). *Managerial Dilemmas*. Cambridge, UK: Cambridge University Press.

Nathan, L. & Myatt, L. (1998). A journey toward autonomy. *Phi Delta Kappan 80* (4), 278-286.

National Center for the Accelerated Schools Project (November, 2000). Accomplishments of accelerated schools. Retrieved from: http://www.sp.uconn.edu/~wwwasp/research/accomplishments.htm

New American Schools (n. d.) How Can We Help? Accelerated Schools Project. Retrieved from: www.naschools.org/contentViewer.asp?highlightID=48&catID=183

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge, UK: Cambridge University Press.

Rowan, B. & Miskel, C. B. (1999). Institutional theory and the study of educational organizations. In J. Murphy and K. S. Louis (eds.) *Handbook of research on educational administration* (pp. 359-383). San Francisco: Jossey-Bass.

Swope, J. (S.J) & Latorre, M. (2000). *Fe y alegria schools in Latin America: Educational communities where the pavement ends*. Santiago: Centro de Investigación y Desarrollo de la Educación (CIDE).

Sociedad de Instrucción Primaria de Santiago (1998). *Reglamento Interno* Santiago: Sociedad de Instrucción Primaria.

Sociedad de Instrucción Primaria de Santiago (1999). *Memoria, 1999* (annual report of the Society). Santiago: Sociedad de Instrucción Primaria.

About the Author

Robert McMeekin

Centro de Investigación y Desarrollo de la Educación
Santiago, Chile
Email: mcmeekin@rdc.cl

Robert McMeekin is retired from the World Bank and lives in Chile. He is an Investigador Asociado with the Centro de Investigación y Desarrollo de la Educación in Santiago and serves as a consultant to the UNESCO Regional Office for Education and other organizations. He holds a doctorate and a Masters in Public Administration from Harvard, a Masters in International Relations from the University of Kentucky and a BA fromYale. A member of the International Society for New Institutional Economics (ISNIE), his book applying NIE concepts to study incentives to improve education has recently appeared.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner

Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

This article has been retrieved 1277 times since May 29, 2003

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

Education Policy Analysis Archives
Volume 11 Number 17 **May 29, 2003** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Secondary Education in Argentina during the 1990s: The Limits of a Comprehensive Reform Effort

Jorge M. Gorostiaga
University of Pittsburgh

Clementina Acedo
University of Pittsburgh

Susana E. Xifra
International Institute of Educational Planning-UNESCO
Buenos Aires, Argentina

Citation: Gorostiaga, J. Acedo, C., Xifra, S. (May 29, 2003). Secondary education in Argentina during the 1990s: The limits of a comprehensive reform effort. *Education Policy Analysis Archives, 11*(17). Retrieved [date] from http://epaa.asu.edu/epaa/v11n17/.

Abstract

The reform of secondary education has been a fundamental part of national educational policy in Argentina since the beginning of the 1990s. Along with the decentralization of responsibilities to provinces and a new structure of primary and secondary education, changes have affected the areas of curriculum design, teaching methods, teacher training, school management, and information and evaluation systems. This study describes the main policies on secondary education implemented during the last decade, including their objectives and rationales. Focusing on how the reform can be seen to relate to issues of access, quality and equity, the study presents an analysis of its implementation, and discusses some of its effects. We argue that political, economic and technical factors as well as the strategies chosen by the national government resulted in a limited implementation, and we highlight the need for considering more focused reform strategies, alternative models of teacher training, and a more active involvement of teachers. (Note 1)

A. Introduction

This article examines the reform of secondary education in Argentina during the 1990s. (Note 2) It provides a brief description of the main elements of the reform, as well as a discussion of the implementation strategies and problems. We highlight different views about the effects of the new policies, particularly as they relate to two of the main objectives of the reform: increasing access for students from economically disadvantaged families and improving the quality of academic and technical secondary education. (Note 3) This study is based on an extensive review of official documents and secondary sources as well as on information provided by key informants. (Note 4)

The reform of secondary education, which involves the redefinition of its structure, its educational functions, and its role in contributing to national development and social integration, is gaining high priority in Latin America (Braslavsky, 2001; Wolff and Castro, 2000). Although there are important differences among Latin American countries, in general, "the region is deficient in the availability of places in secondary education, as well as in the provision of quality, efficiency and equity. The structure and often the content of secondary education is outdated" (Wolff and Castro, 2000, p. 5).

In the case of Argentina, a comprehensive reform of the educational system initiated at the beginning of the 1990s proposed profound changes in the organization of secondary education, attempting a combination of "reculturing" (changing norms, habits, skills and beliefs) and "restructuring" (changing formal structures) processes both at the school and at central levels (see Fullan, 1993). "Restructuring" took place through the decentralization of the administration of national secondary schools to provinces; the establishment of a new structure that extends compulsory education from seven to ten years and creates a common upper secondary school with different orientations; the establishment of a new mechanism of curriculum design; and the development of evaluation and information systems. Elements more related to "reculturing" included the re-training of teachers aiming at upgrading subject knowledge and teaching methods; and the development of a new model of management that encourages school autonomy.

As Levin (2001, p. 6) points out, "The task of the analyst… is to consider the ways in which policies are driven by a particular logic or ideology, but also the ways in which they are shaped by other factors—historical, cultural, institutional, and political—that are far less predictable." In this article we argue that the case at hand represents an ambitious effort of comprehensive reform, complicated by the dynamics of a federal system. This effort encountered important barriers in political, economic and technical factors, but also as a result of a reform strategy that neglected broad discussion and participation, particularly from teachers. Such barriers shaped the reform policies in various ways, as is particularly illustrated by the different provincial responses and the teachers' reactions.

B. The Argentine Education System, the Context, and the Rationales for Secondary Education Reform

Since the beginning of the 20th century, the structure of the Argentine school system was typically organized around a compulsory primary education of seven years beginning at the age of six, and secondary education system with schools usually offering five years of instruction. In 1905, a law authorized the federal government to create and administer its own schools within the provinces, establishing a national system besides the various provincial systems of primary and secondary schools. Initially, secondary education had the objectives of preparing students for teaching at the primary level or for entering the University, but through the years, new types of secondary schools were created, most notably technical schools with a vocational orientation. (Note 5)

During the 1980s and with the reestablishment of democratic rule – which brought about greater responsiveness to the demands for access through the elimination of entrance exams and the construction of new buildings for secondary schools – a remarkable expansion of secondary education took place. (Note 6) The secondary net enrollment rate went from 33.4% in 1980 to 53.5% in 1991. (Note 7) Regional differences, however, remained high: in 1991, the city of Buenos Aires showed a *net* enrollment of around 72% while the figure for a province like Chaco (in the less-developed northeast region) was approximately 38% (Tiramonti, 1995).

The reestablishment of democratic political institutions also allowed for a more open and broader-based discussion of the structure and content of education. With the Pedagogic Congress (Congreso Pedagógico Nacional, 1986-88), which involved teachers, parents, community members, students and representatives of different organizations, "an effort began to build consensus and to seek out or create new policies and action strategies" (Braslavsky, 1998, p. 299). As a result of this effort, the government began to work on the design of a new law of education.

During the 1990s Argentina experienced a profound reform of the educational sector as part of a general restructuring of the state, as well as of attempts to 'modernize' the country and adapt to international economic competition. Carlos Menem's administration (1989-1999) carried out a deep neoliberal economic reform, including the opening of markets to international trade, the privatization of state-owned companies, and the deregulation of economic activities. The Argentine economy showed high growth rates during the early 1990s (averaging more than 8% annual growth between 1991 and 1994), but the decade ended with a situation of recession and uncertainty. At the same time, "poverty levels stubbornly stayed high despite rapid economic growth… [and] rising income inequality and high unemployment, especially for the unskilled, [indicated] that the benefits of growth [were] not widely shared" (World Bank, 2000a, p. 3). (Note 8) The influence of international lending organizations has been very important in the process of economic restructuring, and the World Bank and the Inter-American Development Bank have participated in the financing of educational reform.

Educational reform was supposed to introduce changes in a public school system considered to be in crisis. (Note 9) There was a general agreement in the need for organizational and instructional innovations to improve quality and to provide students socially meaningful knowledge (Braslavsky and Tiramonti, 1990; Frigerio, 1995). National and provincial ministries of education appeared to lack the capacities to govern the system and affect the work of schools (Braslavsky, 1998). The fragmentation of the system implied great inequalities, as schools varied in the quality and type of services they offered to different groups of students. During the 1980s, the combination of increases in enrollments and lack of investment in the educational system resulted in overcrowded schools and classrooms and accentuated the sense of crisis.

The *gross* enrollment rate became more stable in the 1990s after the remarkable growth of the early 1980s (see Table 1). By the end of the 1990s, the secondary graduation rate was around 52% (World Bank, 2000). The repetition rate was 9% for 1997, while the dropout rate for 1996 was 42% (Experton, 1999). It is also important to note, again, that there have been are significant variations among provinces, and that repetition and dropout rates have been particularly high among students from low-income families.

Table 1
Secondary Education Gross Enrollment Ratios in Argentina (1970-97)

Year	Gross Enrollment Ratio
1970	44.4
1975	53.8
1980	56.2
1985	70.2
1990	71.1
1995	72.7
1997	73.3

Source: UNESCO Indicators, in www.unesco.org/education/information.

Some of the main problems affecting secondary education at the beginning of the 1990s were: the high dropout and repetition rates (particularly during the first and second years of secondary school), outdated curriculum and instructional methods, and the limited enrollment of students of low socio-economic status (Braslavsky, 1999; World Bank, 1995). Technical education especially suffered a low level of investment in educational resources as well as a lack of relevance of educational contents (Experton, 1999). The coexistence in the provinces of a national and provincial system of secondary education implied inefficiencies and conflicts regarding aspects like financing and the transfer of students from one system to another. In addition, the administration of the *national* secondary system was seen as too centralized as well as lacking effective supervision over schools (Morduchowicz, 1999).

C. Educational Reform during the 1990s

The two main steps in the reform of education were the transfer of all national secondary schools and teacher training institutes to the provinces in 1992-93 (Law No. 24049) and the enactment in 1993 of the Federal Law of Education (Law No. 24195), the first general law of education in the history of Argentina. (Note 10) The Federal Law of Education (along with the Law of Higher Education enacted in 1995) spelled out a comprehensive reform of the whole educational system.

At the beginning of the 1990s, around 60% of public secondary schools were under national administration. However, there were significant differences between provinces. In some provinces (e.g., La Pampa, San Juan, Tucumán), secondary education was mainly provided by private and public national schools, and the provincial sector was very small. In provinces like Chaco, Rio Negro and Santa Cruz, less than 15% of secondary schools depended on the national government (Tiramonti and Braslavsky, 1995, pp. 61-62).

The transfer of all national secondary schools and post-secondary institutions to the provinces in 1992 – completing the process initiated by the military government in 1978 with the transfer of national primary schools – was mainly driven by financial reasons (Filmus, 1998; Senén González and Arango, 1997). (Note 11) Provinces agreed to receive the national secondary and post-secondary systems, but were not given specific resources to face the economic effort that the transfer represented. The national government argued that from 1991 significant increases in tax collection implied more resources for the provinces, which allowed them to finance the administration of the transferred schools (Senén González and Arango, 1997).

In addition, many provincial administrations lacked the necessary technical expertise and resources to manage the new system (García de Fanelli, 1997; Puiggrós, 1997). In some provinces, the decentralization was followed by an actual reduction of teachers' salaries in the context of a fiscal crisis accelerated by the burden that the transfer of schools implied (Filmus, 1998; Senén González, 1997).

The Federal Law of Education (1993) established the responsibilities for each level of government as well as the coordination mechanisms. The national Ministry of Culture and Education is now responsible for evaluating and monitoring the educational system, ensuring adherence to national policies, providing financial and technical assistance to improve the quality and equity of the system, and developing a federal management information system. The 23 provinces and the city of Buenos Aires (federal district) have the responsibilities of funding, administration, and management of schools, including the hiring and training of teachers. (Note 12) The Federal Council on Culture and Education (*Consejo Federal de Cultura y Educación*) is the organization in which the national policy and its implementation are discussed with the provinces. The national and provincial ministries of education are part of this organization.

The new model that emerges is one of a stronger center with few (but strategic) responsibilities (e.g., the planning of general policies and evaluation of the system). Another important element of the law was that it established an increase of the resources for the financing of education from national and provincial governments: educational budgets were to be doubled gradually with an annual increase of 20%.

According to Decibe (1998, pp. 3-4), who was one of the ministers of education during the reform, the decentralization of secondary schools and the new division of responsibilities aimed at strengthening federalism and local government institutions, including the schools. In this process, she argues, the national government should keep the capacity to lead policies, given the links between education and the possibilities of economic and social growth.

What follows is a description of the main components of the reform that were prompted by the Federal Law of Education, including: a new structure of the educational system, new systems of curriculum design, teacher training, information and evaluation, a new model of school management, and the compensatory programs.

1.The New Structure of the System

The Federal Law of Education (1993) created a new academic structure which starts with an initial level up to the age of five, continues with a nine-year stage of *Educación General Básica* (Basic General Education), or EGB, plus three years of *Polimodal* Education, and ends with the higher-education level (comprising universities and tertiary institutions). Compulsory schooling comprises a ten-year period: the last year of the initial level (5-year-old children) and a nine-year stage corresponding to the EGB, divided in 3 cycles of three years each. Secondary or middle education can be considered to include now the EGB3 (7th, 8th and 9th grades or years) and *Polimodal* (or Multimodal) cycles (see Figure 1).

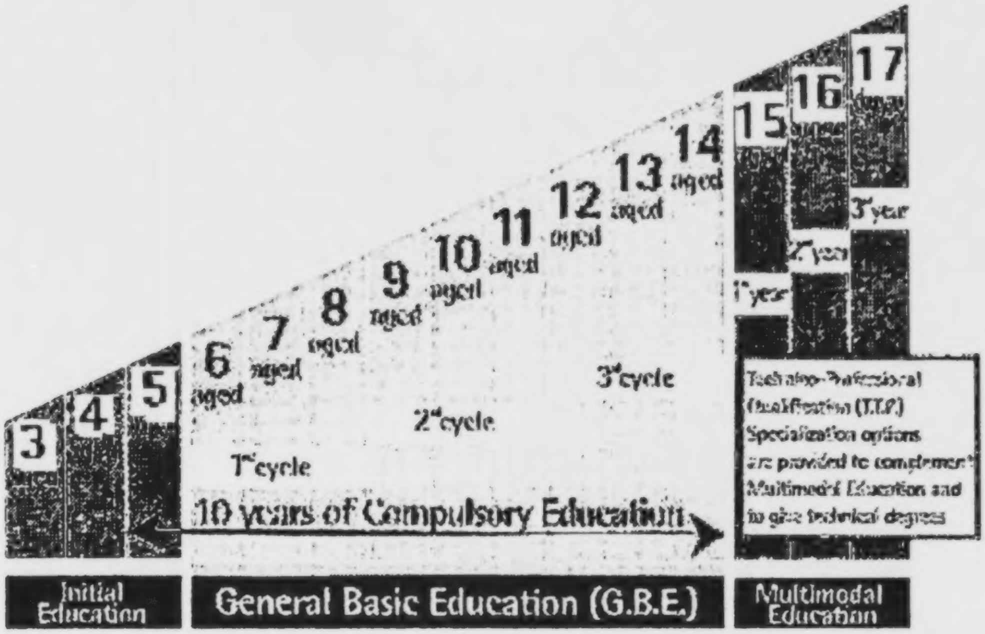

3
4
5
6 aged
7 aged
8 aged
9 aged
10 aged
11 aged
12 aged
13 aged
14 aged
15
16
17
1st year
2nd year
3rd year
1st cycle
2nd cycle
3rd cycle
10 years of Compulsory Education
Technico-Professional Qualification (T.T.P.) Specialization options are provided to complement Multimodal Education and to give technical degrees
Initial Education
General Basic Education (G.B.E.)
Multimodal Education

Figure 1. The new structure of the Argentine education system
Source: National Ministry of Culture and Education (Ministerio de Cultura y Educación, 1997)

In the national reform strategy, the creation of the Third Cycle was meant to address what many considered an abrupt change for students in the transition from primary to secondary school (e.g., Braslavsky, 1999), and to respond to the specific needs of young teenagers (Ministerio de Educación, 2000a, p. 6). In addition, the implementation of both the Third Cycle and the *Polimodal* should have involved strategies for improving equity and the quality of instruction for low-income students.

The *Polimodal* lasts 3 years and offers a common general education (comprising at least 50% of the instructional time) (Note 13) along with five different orientations or modalities (around 30% of instructional time): (Note 14) humanities and social sciences; natural sciences; economics and administration/management; production of goods and services; and communication, arts and design. Each school can offer one or more modalities and students have to choose one modality. The *Polimodal* should address with equal weight three different aspects of the education of students: preparation for citizenship and ethics, preparation for work, and preparation for further studies (Consejo Federal de Cultura y Educación, 1996).

Professional technical training is offered through *Trayectos Técnicos Profesionales* (TTPs), or Technical Professional Qualifications. These courses of study can be followed with or after the Polimodal, and grant the title of "technician" in different areas, including not only those linked with construction, agriculture and industry sectors, but also others like public health, computer science, tourism, etc. (Ministerio de Cultura y Educación, 1997). Provinces should decide how to organize the supply of different orientations and different technical specializations, taking into account local needs and school institutional capacities.

BEST COPY AVAILABLE

According to Braslavsky (1999), the organization of the *Polimodal* and the technical specializations respond to the need for training students in a way that enable them both to find a first job and to adapt to a constant changing job market. At the same time, they offer a common educational experience along with enough flexibility to integrate the content of training to the characteristics of local communities and contexts. In this way, the new organization of secondary school, including the technical component, seems to aim at finding a balance between a general education that enables students to advance to higher levels of instruction and a technical-professional training that prepares for specific jobs.

In 1995, the National Ministry created the Institute of Technological Education (INET), which has organized the implementation of technical specializations by adapting the centrally defined curricular frameworks to local realities, integrating them to the *Polimodal*, and identifying training needs for teachers in technical areas (Experton, 1999, p. 27). One of the main functions of the INET is to respond to the needs of the productive sector. It is argued that a new relationship of education-employment is being established by the creation of more effective mechanisms of communication and debate among teachers, educational officers, and representatives of workers and businessmen (Ministerio de Cultura y Educación, 1997).

The Federal Council established that the new structure should be completely in place by the year 2000, allowing each province to decide on how to implement it (e.g., more or less gradual, use of pilot schools, etc.). (Note 15) The EGB3 has been gradually implemented in almost all provinces since 1996. Regarding the *Polimodal*, the provinces of Córdoba and Buenos Aires started its full implementation in 1999, while most provinces have a slower pace.

2. Curricular Reform

The curricular reform has been another important initiative encouraged by the national government since 1994. The national Ministry and the Federal Council worked out a general framework for all of the elementary and high schools in the country, aiming at giving coherence to the educational system. Traditionally, provinces presented a great heterogeneity of educational programs (Gvirtz, 1995) and, for each type of secondary education, there was a common curriculum which was adapted further at the school level, resulting in an "anarchic diversity" (Braslavsky, 1999, p. 84).

There are three levels in the design of the new curricula. At the more general level, the Federal Council sets the general objectives and guidelines (Common Basic Contents or "contenidos básicos"). (Note 16) At the provincial level – including the city of Buenos Aires – more specific objectives and guidelines are developed which take into account the reality of each province. The final design of the curriculum at the school level involves principals and teachers making decisions on content and instructional approaches. The Federal Council has stated that the school is the fundamental unit for the specification of the educational project, so the nation and the provinces should leave to the schools the responsibility of developing a curriculum that responds to the local realities and needs ("Criterios para la planificación de Diseños Curriculares Compatibles en las Provincias y la MCBA, res. 37/94," in Ministerio de Cultura y Educación, 1996b). At the *Polimodal* level, in addition to the adaptation of provincial curriculum designs, 20% of the content should be determined by the school (Consejo Federal de Cultura y Educación, 1996).

In addition to the Common Basic Contents, in 1998, when provinces like Buenos Aires and Córdoba were already implementing the third cycle of the EGB, the Federal Council agreed to establish a common curricular structure for the EGB3 with nine "curricular spaces:" Spanish, mathematics, social sciences, natural sciences, foreign language, arts education, technology, ethics and citizenship, and physical education. A similar agreement was reached for the *Polimodal* and the technical specializations, establishing the five modalities mentioned above, and the possibility of a variety of technical-professional training courses.

3. Teacher Training

The design of new curricula implied the need for the re-training of teachers and principals. Teachers were considered to have a very low degree of both pedagogical skills and subject knowledge. The Federal Network of Teacher Training was designed to prepare teachers for the curricular changes, update their skills in subject matter, and improve their use of educational materials and computers. Specific training – which took place outside schools – was offered for teachers at the EGB3 and at the *Polimodal* cycles, while encouraging university education for teachers at the *Polimodal* cycle. An innovative aspect of the network was that principals, also, received specific training.

Both Teacher Training Institutes (either under provincial or private administration) and universities are considered to be part of the new training system, under the supervision of the provincial ministries. The curricula of Teacher Training Institutes were also redesigned with three levels of specificity: national, provincial, and institutional. The new system was supposed to train future teachers as well as retrain all of the teachers already working at schools (World Bank, 1995; Decibe, 1998).

4. Education Information and Student Evaluation Systems

A National System of Educational Quality Assessment has been created in the last years. Since 1993, a national evaluation has been done yearly using two kinds of instruments: 1) multiple choice tests to evaluate students' performance in different disciplines, and 2) questionnaires to school administrators, teachers, and families to gather information about institutional management, classroom practices, and students' background, attitudes, and habits of study (Decibe, 1998).

Multiple-choice tests have been administered to samples of students at primary and secondary levels, and the results have been publicized. From 1997, all students finishing the secondary level are evaluated in the areas of Spanish and mathematics. The tests have shown that students are not performing well, that private schools do better than public, and that urban schools do better than rural. The results have increased awareness about the low quality of instruction at schools and provide some legitimization for the reforms.

The National Network of Educational Information has helped advancing in the development of (and coordination among) provincial units of educational information as well as in the use of information for decision-making. The first national census of teachers and educational institutions took place in 1994 (Decibe, 1998).

5. A New Model of School Management

The national government encouraged from the mid 1990s a new model of school through the *Condiciones Básicas Institucionales* (Institutional Basic Conditions). The institutional conditions were proposed from the national level to the provinces as guidelines for the management of all schools in the country (Ministerio de Cultura y Educación, 1996c) and were related to the objectives of consolidating the school as the basic unit of management and strengthening its autonomy (Consejo Federal de Cultura y Educación, 1994).

The main idea is that the application of new curricular guidelines requires changes in the management and organization of schools in order to be effective and to facilitate instructional innovations. Changes in the management dimension have the objectives of encouraging the participation of the school actors and facilitating the organizational changes that would result in the transformation of instructional practices. It is argued that each school should make its own decisions on the use of the time (e.g., replacing the traditional classroom hour by different units of time), the use of space (e.g., classrooms in which students are allowed to move among "corners" with different activities), and grouping practices (e.g., more than one teacher working with students in a jointly planned activity; students of different ages grouped together for a special activity) (Ministerio de Cultura y Educación, 1996c).

The Institutional Educational Project (IEP) became the main tool for the new model of school. It is a document produced by each school elaborating and adapting the Provincial Curricular Design and deciding how to deliver the curriculum. Each school is invited to build its own organizational and management structure based on its needs, its reality, and the people who are part of the school. The IEP is the mechanism that should allow the school to mediate between the general educational guidelines (at national and provincial levels) and the local context. The official proposals stress the importance of teamwork, involving all of the school actors and the local community working on defining the mission and objectives of schools and how these are achieved both at the institutional and classroom levels. The instructional practices should be the "result of an action which is deliberately and consciously planned by the institutional [school] actors, within the frame of the national and provincial political guidelines" (Ministerio de Cultura y Educación, 1996c, p. 33). The central element of the IEP is the school curricular project in which teachers and principals are given central participation.

6. Compensatory Programs

The Federal Law established a compensatory role for the national government, which translated into the Social Plan for Education (*Plan Social Educativo*), designed to provide federal funds for improving facilities, computers and textbooks directly to the poorest primary and secondary schools in the country. The Social Plan reached more than 17,000 schools and about 3.6 million students during the period between 1993 and 1998; in the Northeast region, for example, it covered nearly 87% of all schools (Morduchowicz, 1999).

The new emphasis on compensatory programs can be seen as responding to a social context of increasing fragmentation and marginalization (see Minujin and Kessler, 1995). The government recognized that social segmentation and impoverishment translate into schools with different levels of resources serving different groups of students. The aim of one of the programs, for example, is "to provide poorer children with the same (pedagogic) resources as the most privileged ones" (Ministerio de Cultura y Educación, 1997, p. 44).

The different programs of the Social Plan included the construction of new classrooms for the EGB3 cycle, the provision of scholarships for students from low-income families to help them attend the EGB3 and *Polimodal* cycles, and the *Proyecto Tercer Ciclo de la Educación General Básica Rural*, which since 1996 provided technical and financial assistance for the implementation of the EGB3 in rural schools. In 1999, this specific project covered about 24% of rural schools in the country, distributed in 21 out of 24 provinces (Gozman and Jacinto, 1999).

D. Implementation Strategies and Problems

In general, the implementation of the reform has been characterized by a new form of relationships between the national and provincial ministries of education, the support of international organizations (like the World Bank, IDB and UNESCO), and the opposition of teachers' unions.

1. The Role of International Organizations

During the period between 1995 and 1999, the loans from IDB and the World Bank for the financing of reforms represented 2% of global educational expenditures (at the national and provincial levels). During that period, the World Bank invested US$410 million for the Decentralization and Improvement of Secondary Education Project, while IDB provided US$600 million for the Educational Reform and Investment Program (Decibe, 1998). The World Bank project focused on the EGB3 and *Polimodal* cycles. It supported: the strengthening and modernization of administrative, evaluation and planning capacities and systems at the provincial level; curriculum development; provision of educational resources (textbooks, libraries, computer and science laboratories, etc.); school expansion and rehabilitation; and the development of school-based improvement projects (Institutional Educational Projects) (World Bank, 1995; 2000). The IDB program focused on the EGB and teacher training.

According to Tiramonti (1996, pp. 15-16), international organizations like UNESCO and the World Bank defined the general agenda of educational reform, but the Argentine government had some possibilities of negotiating how to use the loans received. Cocorda (1999), on the other hand, holds that the government tended to invest in the areas prioritized by the World Bank and IDB without its own diagnostic of the educational sector.

2. The Relationship Between the Central Government and the Provinces

In the implementation of the Federal Law and related changes, the national government followed two strategies in the period between 1994 and 1999 (Senén González, 2000): 1) working with the provincial MOEs to build consensus and legitimacy through the Federal Council on Education; and 2) working directly with schools through specific programs. Another characteristic of the manner in which the national government guided the reform was to initiate all of its elements simultaneously, attempting to complete many of them in a relatively short time (Experton, 1999, p. 7).

As was shown in the previous section, the national government played a prominent role in designing educational policy and programs, some of which were supposed to support and coordinate provincial efforts in the main aspects of the reform. In the new scenario established by the Federal Law of Education, the national policy and its implementation were to be discussed with the provinces at the meetings of the Federal Council on Culture and Education (*Consejo Federal de Cultura y Educación*). The Federal Council advanced in the definition of some of the main elements of the reform like the new structure of the system, the new system of educator training, the curricular reform, and the development of the evaluation system.

Apparently, the Federal Council did not have an active role in the production of proposals, but discussed the initiatives and documents produced by the National Ministry of Culture and Education. The decisions of the Federal Council are not compulsory for the provinces, but some argue that the provinces' financial dependence on the national government and lack of expert knowledge make them respect those decisions (Tiramonti, 1996). (Note 17)

As mentioned before, an important element of the law was that it established that national and provincial governments should significantly increase the resources for the financing of education. Although the specified goal was not achieved (public educational budgets were supposed to reach 6% of the gross national product (GNP)), educational expenditures did grow in the context of a general expansion of public expenditures. (Note 18)

The Federal Pact on Education (*Pacto Federal Educativo*) was signed in 1994 by all provincial governors and the President, and it established a commitment from the national government to invest US$3 billion during a five-year period to fund provincial needs in the areas of educational infrastructure and equipment, and teacher training. Provinces committed to implementing the Federal Law and to investing in the reforms (Consejo Federal de Cultura y Educación, 1994).

From 1995, when the Mexican financial crisis impacted Argentina, the expenditures expansion decreased (see Table 2). In the period between 1992 and 1997, the national government almost doubled its educational expenditures while provinces (including the city of Buenos Aires) increased their expenditures to about 40% (see Table 3).

The strategy for involving provinces was facilitated by the fact that the same political party (*Peronismo* or *Partido Justicialista*) administered the national and almost all provincial governments during the 1990s.

Table 2
General Educational Expenditures in the Period 1992-1997 (in US$ millions)

1992	US$ 7.898
1993	US$ 9.531
1994	US$10.844
1995	US$11.136
1996*	US$11.302
1997*	US$11.796

* The data for these years are provisional.
Source: D.N.P.G.S.—Secretaría de Programación Económica. Ministerio de Economía y Obras y Servicios Públicos, Argentina.

Table 3
Educational Expenditures (in US$ millions) for the Period 1992-1997, by Level of Government

Year	National	Provinces and City of Buenos Aires	Municipalities	Total
1992	1.362	6.057	159	7.578
1993	1.691	7.135	190	9.016
1994	1.985	8.066	204	10.255
1995	2.086	8.273	194	10.553
1996*	2.213	8.369	196	10.778
1997*	2.556	8.490	200	11.246

* The data for these years are provisional.
Source: D.N.P.G.S. - Secretaría de Programación Económica. Ministerio de Economía y Obras y Servicios Públicos.

The only jurisdiction that strongly resisted the national proposals was the city of Buenos Aires, which enjoys a high level of resources and where Peronism has always been a minority in electoral terms. The provinces of Buenos Aires and Córdoba, with high level of resources, also showed some independence from the national government (Senén González, 2000).

In general, provinces appeared to accept the national initiatives because of the additional funds these initiatives represented in a context of provincial fiscal crisis (Senén González, 2000). However, many provinces had an inadequate level of financial and technical resources to implement various aspects of the reform proposed at the central level (Experton, 1999; García de Fanelli, 1997). Some analysts argue that the national government between 1993 and 1999 failed to support technically the provincial MOEs – which presented differing needs and capacities – for an effective implementation of the new policies (Senén González, 2000; Roggi, 2001).

3. Implementation at the Provincial Level

Provinces adopted different responses to national policies and projects, according to their financial and technical resources as well as local politics. An example of the different provincial strategies for the implementation of educational changes was the reform of the structure of primary and secondary education. Provinces like Buenos Aires and Córdoba designed their own policies with regard to the localization and the role of the EGB3 cycle, which includes the last year of the former primary and the first two years of the former secondary cycle. The province of Buenos Aires decided to implement an "institutional model" of EGB that includes the three cycles, from the 1st to the 9th year or grade. (Note 19) By contrast, Córdoba decided that the EGB3 and the *Polimodal* should constitute two cycles of the new Middle Schools. Most provincial governments seemed to choose a mixed model, where the third cycle of the EGB is placed in new buildings or in schools offering either the other two cycles of EGB or the *Polimodal*, depending on the availability of space at the buildings, their geographical location, and in negotiation with different actors (community, supervisors, school principals, local teachers' union, parents) (Ministerio de Educación, 2000b). In the cases of the city of Buenos Aires and the province of Río Negro, the new structure has not been implemented at all (Ministerio de Educación, 2000d), while many other provinces have recently started the introduction of the new cycles, involving a limited number of schools (Aguerrondo et al., 2000). (Note 20)

Provinces have also showed significant variations with regard to curricular aspects. In terms of the number of subjects, for example, 13 subjects are offered in the EGB3 cycle in Cordoba, while in the province of Buenos Aires there are just seven "curricular spaces." At the same time, most provinces seem to be increasing the number of hours at least for the academic or traditional curricular areas (language, math, social sciences and natural sciences), but in some cases, like in the province of Salta, they do not respect the minimum established (Ministerio de Educación, 2000c, p. 7).

For the EGB3 cycle, which implies the extension of compulsory education to the 8th and 9th years, provinces have faced financial problems as well as challenges in the availability and relocation of teachers. In some provinces, decisions about who should teach at this level (primary *maestros* vs. secondary *profesores*) and with what kind of retraining were not carefully planned. Many appointed teachers did not have the necessary training, and many posts were not covered at all (Ministerio de Educación, 2000b). In most cases, both primary and secondary teachers have been included in the EGB3 cycle, but this has resulted in tensions between the two groups and the particular ways in which they conceive the teaching and learning processes (Ministerio de Educación, 2000e, p. 18). Another problem has been the lack of training for teachers who would hold new positions as "coordinator" of the cycle or "tutors" (new positions created with the EGB3) (ibid, p. 12). There are also questions about the capacities of principals to deal with the new structure of basic education (Krichesky and Capellacci, 1999).

The national initiative for the retraining of teachers appeared as one of the most difficult reform elements to implement. According to the National Confederation of Teachers' Unions and based on a survey of teachers' opinions, teachers have found difficulties attending training courses, mainly because of financial and time limitations (CTERA, 2001). In some cases, the content and quality of teacher training courses have not satisfied teachers (Casassus, et al., 1998).

4. The Relationship Between Central Government and the Schools and the Movement toward School Autonomy

The Social Plan for Education (*Plan Social Educativo*), previously described, can be considered as an example of the direct relationship that the national government established with some schools. Another example is New School (*Nueva Escuela)*, a program implemented from 1994 (reaching 1100 schools in 1995, though most of them in an indirect way) (Tiramonti, 1996). The New School program attempted to facilitate the development of the new and more autonomous model of school management. Similar to the mechanism employed in the Social Plan, each province selected a coordinator and a group of schools to participate in the program; the national government provided funds and technical assistance for carrying out the school projects. (Note 21)

With regard to the implementation of the new model of school management, it has been evaluated that decentralization at the school level was more rhetorical than real (Experton, 1999, p. 59), while the design of the IEPs at schools was done with limited participation of teachers and local communities and with a lack of measurable and clear objectives (Experton, 1999, p. 28).

The differences among schools in terms of cultural and material resources are a big challenge for moving toward more autonomous forms of organization and management if autonomy is expected to improve the quality of education for all. In the words of somebody who was actively involved in the analysis and planning of educational reform during the 1990s, "The main risk in the current reconversion of the educational system is that only private schools attended by the higher-income population, and public schools that participate in plans and projects for offering targeted assistance, will be able to gain the kind of autonomy needed to adopt their own proposals for restructuring the system" (Braslavsky, 1998).

5. Teachers' Salaries and Political Resistance

In addition to the limited capacities of provinces, low teachers' salaries and the unions' opposition to reforms have been considered strong barriers for the implementation of the new policies. (Note 22) Teachers' unions seemed to see the reforms implemented during the Menem administrations (1989-1999) as part of a neoliberal policy that attacked public schools and tried to weaken unions' power (CTERA, 1997). In the province of Buenos Aires, for example, the implementation of the new structure has been sometimes perceived by teachers as a threat to their working conditions and their identity (Krichesky and Capellacci, 1999), and they have had very limited possibilities of influencing the process of educational reform (Casassus et al., 1998). According to the cited national survey produced by CTERA, 69% of teachers report that they did not participate in the elaboration of the new curriculum, and 65% think that it hasn't had a positive effect on teaching (CTERA, 2001).

With regard to salaries, in 1996, teachers at the secondary level earned 60% less than in 1980 (Carnoy and Castro, 1996). Llach, et al (1999), show that teachers' average initial salary is less than 40% of the average salary in the formal economy. The national government started to explicitly recognize the problem of salaries in 1998, after a year of teachers' manifestations, with a law that provided a small increase for all teachers in the system (La Nación, 7/9/99).

The resistance of teachers' unions and of important sectors of the main opposite political parties to significant aspects of the reform put into question the continuity of the Federal Law when, by the end of the 1990s, a party change in the national administration was imminent. Representatives of the political opposition pointed out that the national government had not attempted to achieve a broad agreement about the direction of the reforms, had not provided the necessary funds for carrying out the reform, and had concentrated resources centrally while leaving provinces with the responsibility of implementation (Delich, 1999; Puiggrós, 2000; Stubrin, 1999). (Note 23)

Overall, the reform advanced in aspects like curriculum and student evaluation, (Note 24) and was accompanied by a major effort in upgrading facilities and providing resources for some of the schools in the poorest areas of the country. On the other hand, it faced many problems in the area of teacher training, made little progress in developing a new model of a more autonomous school with participation of teachers and local community, and resulted in an uneven implementation of the new structure.

Many of the proposed changes may require a sustained, long-term effort, more feasible if it is based on social consensus and the involvement of a wide range of actors. As Fullan (1993, p. 49) points out, "the hardest core to crack is the learning core—changes in instructional practices and in the culture of teaching toward greater collaborative relationships among students, teachers and other potential partners" as well as, we would add, among educational bureaucrats, teachers, principals and supervisors. "Stated differently, to restructure is not to reculture… Changing formal structures is not the same as changing norms, habits, skills and beliefs" (Ibid.). In the case of Argentina, there are indications that while the process of restructuring has faced significant barriers, it is at the level of reculturing where reform strategies are more in need of refinement, and of securing the support of educational actors as well as the different levels of government.

E. The Effects of the Reforms: Quantitative Expansion and Quality Improvement?

It is difficult to evaluate the effects of secondary education reform since its implementation is a recent process. Some authors hold that the "primarization" of the old 1st and 2nd years of secondary education (now the 8th and 9th years of the EGB) has negative pedagogic effects since it involves a uniform approach that does not respond to the specific needs and problems of teenagers (see Puiggrós, 1997). In addition, it is argued that the different structures of secondary education within and among provinces result in a fragmented system (FIEL, 2000; Puiggrós, 2000).

A document produced by FIEL (Fundación de Investigaciones Económicas Latinoamericanas), a research institution linked to business groups, regrets the neglect of technical education, states that "the extension of a generalist academic education [through the EGB] does not appear as the best strategy for student retention" (p. 105), and concludes that the reform of the educational structure has been "costly, harmful in some aspects, and not necessary in other aspects" (FIEL, 2000, p. 110, authors' translation). According to this document, compulsory education could have been extended without changing the structure, while technical education should be offered at the EGB3 cycle, which would make schooling more attractive and useful for students from low economic status.

Teachers' unions also tend to criticize the change of structure and its pedagogic effects. Some of the critiques that teachers' unions make to changes affecting secondary education are reflected in the following quotes from a document produced by one of the most representative groups of the province of Buenos Aires:

> [The EGB] deepened the fragmentation of the educational system and caused fractures in the school organization... This discourse [of the reform] translated into lack of consideration of the characteristics of teenagers, of institutional cultures, of work conditions, and of teacher training... One of the most harmful effects of the reform is the disarticulation and disappearance of middle, technical and agrarian schools (SUTEBA, 1999, pp. 4 and 6, authors' translation).

By contrast, other analyses argue that learning conditions have improved through the updating of curriculum content and teaching methods, as well as through the upgrading of buildings and equipment (Braslavsky, 1999). National and provincial governments that implemented the reform during the 1990s and the international organizations that supported it tend to point to increases in enrollment and retention rates as one of the main benefits of the changes introduced (see, for example, Decibe, 1998). (Note 25) The national government also claims that the evaluations of students' learning show improvements in the performance of students, particularly in the poorest regions of the country, and that such improvements can be related to the educational compensatory actions of the federal government (Decibe, 1998, p. 24).

In relation to the effects of the reform on enrollments, some studies have been produced for the province of Buenos Aires, one of the most advanced in implementing the reform, and the most important in terms of the number of students, teachers and buildings. Most studies argue that the reform has been effective in increasing enrollment at the secondary level, but they point to the enduring problems of repetition and ineffective retention. Morduchowicz (1999) reports that in 1997, the number of students attending 8th year (previously the 1st year of secondary) increased 14%, while Casassus et al. (1998) show that the retention rate for the same grade and year increased about 17%. Suárez and Balduzzi (1998), on the other hand, point to the relative value of enrollment increases. They argue that the incorporation of many students was formal rather than real because of the lack of conditions for student learning and for assuring their permanence in the system. They show that only 27% of the students enrolled in 8th year during 1997 achieved the minimum learning requirements, while around 11% of the students dropped out. At the same time, according to their data, only 16% of students were absent from class less than three days per month on average. Krichesky and Capellacci (1999) also show some increases in the number of students attending the 8th and 9th years, while suggesting that schools do not employ effective strategies for the retention of students, particularly those of low SES.

Other studies about the implementation of the EGB3 in different contexts also point to improvement in enrollment rates. In the case of the province of La Pampa, the trends toward increasing enrollment and decreasing dropout rates were accelerated by the implementation of the EGB3 (Duschatzky, et al, 1999). Increases in enrollment and retention rates were also found for the rural schools included in the special project of the Social Plan for the implementation of the EGB3 (Gozman and Jacinto, 1999). (Note 26)

On the other hand, a document of the National MOE (Ministerio de Educación, 2000b) suggests that in the case of the province of Córdoba, the emphasis on improving quality that has been the rationale for the creation of the new Middle Schools comprising the EGB3 and *Polimodal* cycles, might contribute to an earlier exclusion of some students from the educational system; disadvantaged students seem to be dropping out after the 6th grade (when they complete the second cycle of the EGB), instead of 7th grade as was the case with the former primary.

With regard to the improvement of the quality of education, the lack of training of teachers and principals to deal with new responsibilities (particularly the management of the nine-year basic education structure and its three cycles) and curricular changes at the EGB3 cycle seems to imply a major barrier. Krichesky and Cappellacci (1999) suggest that in schools serving students of low SES, teachers and principals see the focus of the EGB3 on providing social and psychological support to students rather than on facilitating learning. Similar constraints seem to affect the *Polimodal* cycle. For the province of Buenos Aires, it is stated that "At the Polimodal level the urgency of improving teacher training and support services is even greater (than at the EGB), since the new streams ("modalidades") require different teaching methodologies and academic skills. The poorest schools encounter most of the problems to implement the new regime, and are most urgently in need of assistance" (World Bank, 2000, p. 6).

The implementation of the EGB3 (from 7th to 9th grade) seems to involve a trade-off between coverage and quality. The linking of this cycle with the previous cycles of general basic education (not only in terms of physical space but also of pedagogic practices) appears to provide more effective integration of students from low-income groups who tended to dropout of school after completion of the former primary level, but it involves the risk of "primarizing" the cycle with a decrease in the quality of teaching. On the other hand, when the EGB3 is integrated into the *Polimodal*, it appears as offering a higher quality education but with less possibilities of retaining "at-risk" students (Ministerio de Educación, 2000a).

F. Conclusion

During the 1990s, Argentina embarked in an ambitious reform of secondary schools and of the general educational system.

While changes have affected almost all aspects of education, there are two elements of the reform that have set the framework for the implementation of other policies. The first one is the decentralization of national secondary schools to provinces and the new division of responsibilities between national and provincial governments. The second element is the new structure of education, which has implied a redefinition of the functions and the content of secondary education.

The reform advanced in defining a new and more coherent system, including the development of evaluation and information systems, and of national curricular guidelines, along with significant investments in infrastructure. Secondary education was restructured, now comprising the third cycle of the compulsory general basic education (EGB3) followed by the *Polimodal* and the technical specializations.

Regarding the origins of these policies, Carnoy (1999) refers to three kinds of responses in the education and training sectors as a reaction to globalization and changes in the world economy:

- Competitiveness-Driven Reforms have as their main goal to raise the productivity of labor and education institutions.
- Finance-Driven Reforms have as their main goal to reduce spending on education, with an ultimate goal of improving the productivity of labor.
- Equity Driven Reforms have the main goal of increasing equality of economic opportunities. The investment in greater equity can be justified when demonstrating that goals of competitiveness are also increased (Carnoy, 1999).

The case of Argentina shows a combination of the three, although the emphasis seemed to be put on the first type. According to Carnoy (1999), competitiveness-driven reforms involve changes in at least the following four categories: decentralization, standards, improved management of educational resources, and improved teacher training. As we have shown, these categories were part of the secondary reform effort in Argentina, and proposed changes were sometimes justified in terms of adapting the educational system to technological changes and the requirements of the global economy (e.g., Ministerio de Cultura y Educación, 1997, 1996b). On the other hand, the transfer of schools to provinces can be characterized as a finance-driven reform, while compensatory programs seem to reflect the equity rationale.

The implementation of the EGB3 has resulted in an increase in enrollment rates by extending the number of years that students from low-income groups stay in the system, but there are questions about the quality of the education offered at this cycle. In addition, the identity of the EGB3 is not clear: continuation of basic education, introduction to upper secondary, or a level that addresses the specific needs and characteristics of young teenagers.

On the other hand, as Castro, et al (2000), suggests, the creation of the *Polimodal* – offering different orientations within a general academic education – may be a legitimate way of addressing "the increasingly flexible labor markets of the new global economy" (p. 14). (Note 27) The combination of the *Polimodal* and the technical specializations appears as a possible way of balancing general academic and more flexible skills formation for technical orientation. There are, however, actors (teachers' unions and business groups) that consider that technical education is being neglected with the possible effect of making secondary education less attractive to students from low- or middle-income groups.

The centralized design of the new policies by a national ministry, which devolved the administration and management functions to provinces in order to focus on the planning and evaluation of the system, has faced problems of implementation derived from the limitation of financial and technical resources at the provincial level. In general, provinces accepted the proposals from the national level, but in many cases were not able (or willing) to provide the necessary funding to carry them out, which resulted in partial implementation, particularly visible in the aspect of the new structure of the educational system. This seems to indicate the need for a more careful design of policies and their implementation, taking into account the particular capacities of different provinces, and for more emphasis on the support to provincial ministries. The financial difficulties for carrying out the various aspects of the reform may be indicative that a more gradual or focused strategy would have been more appropriate.

The analysis suggests that there is also a need for a more careful design of training policies and programs. As Fullan (1993) emphasizes, reform efforts worldwide tend to underinvest in teachers' education. In the case of Argentina, it might make more sense to organize in-service training of teachers and principals at the schools where they work, articulating such training with the design of school institutional projects (particularly the design of the school curricular project and, at the *Polimodal* level, the definition of the 20% of instructional content that is left to the school). There is also a need for specific training to educate students from economically disadvantaged families and developing pedagogic strategies for their retention at the EGB3 and *Polimodal* levels, without disregarding the crucial role that central levels should play through compensatory policies.

At the same time, the Argentine case appears to illustrate the need for involving teachers and their unions in a more active way, and for improving their working conditions, including salaries. (Note 28) It has been recognized that national authorities failed to include non-governmental organizations like teachers' unions in the discussion and negotiation of educational policies, and were not able to create a broad public support for the proposed changes (Braslavsky, 1999).

Although the reform has addressed significant issues related to access, quality and equity, it is not clear at this point in time what elements of the reforms are going to be continued, modified or strengthened in the long term. By the end of the Menem administration, even some of the most critical voices (e.g., FIEL, 2000; Puiggrós, 2000) considered that it would not be wise to revert the reforms given that some provinces had considerably advanced in their implementation. During the De la Rua administration (Dec. 1999-Dec. 2001), the Federal Council of Education decided to give more flexibility to provinces in the implementation of the new structure, while the National MOE began to reformulate aspects like teacher training, compensatory programs, and the evaluation and information systems, giving high priority to the improvement of secondary education. During 2002, in the context of a most severe social and economic crisis, national educational authorities focused efforts on securing funds from the IDB for the continuation of the Scholarships Program. A careful assessment, hopefully through a participatory process, of the policies carried out during the last decade in order to identify how quality and equity aspects can be enhanced in the short and long terms is still due.

Notes

1. This article builds upon research conducted at the Institute for International Studies in Education (University of Pittsburgh) for the Secondary Education Reform Project, which was funded by USAID and directed by Clementina Acedo (see Acedo, 2002). We are grateful for the valuable feedback provided to previous versions of this paper by: Ana Donini (Universidad Nacional de San Martín), Mark Ginsburg (University of Pittsburgh), John Hatch (USAID), Mariano Narodowski and Milagros Nores (Centro de Políticas Educativas, Fundación Gobierno y Sociedad), and MinHo Yeom (University of Pittsburgh).
2. The article focuses on the reform that the Menem administrations (1989-1999) carried out. In December 1999, a center-left coalition replaced the Peronist party in the presidency of Argentina.
3. We focus on *implementation* and, to a lesser extent, on *outcomes*, with some references to the *origins* and to the process of *adoption* of the "policy stages" of reform policies (for a discussion of the stages of educational policies see Levin, 2001).
4. The identification of official documents as well as the contacts with informants were carried out by the third author, based in Buenos Aires, during the first semester of 2001. Secondary sources were identified through literature searches and through conversations with Argentine researchers that the first author held in the context of this and other research projects on Argentine educational reform. Key informants included: Daniel Agostino (Student Scholarships Program, National Ministry of Education); Inés Aguerrondo (International Institute of Educational Planning-Buenos Aires); Marta Andrade de Lago and Javier Rubio (National Center of Educational Information, National Ministry of Education); Verónica Batiuk, Silvia Finocchio, Víctor Meckler, and Daniel Pinkasz (Curriculum Management and Training Program, National Ministry of Education); Guillermo Gozman; Margarita Poggi (Former Director of Educational Planning, City of Buenos Aires); and Cristina Tomassi (Undersecretariat of Basic Education, Province of Buenos Aires). Researchers who have collaborated at different times for the identification of secondary sources include: Myriam Feldfeber (Universidad de Buenos Aires), Mariano Narodowski (Universidad Nacional de Quilmes), Silvia Senén González (Universidad Nacional de General San Martín), and Guillermina Tiramonti (FLACSO, Buenos Aires).
5. In 1986, 75 % of secondary students were enrolled in the general (more academic) tracks (Bachiller and Comercial), around 20 % in technical schools, and around 4 % in the more specific tracks (like agricultural or artistic schools) (Aguerrondo, 1996, p. 110).
6. In December 1983, a democratically elected government took office, an event that would mark the end of more than fifty years of alternation between military and democratic governments.
7. Enrollment rates relate to enrollment in a particular level of education to the age cohort that commonly would be expected to be participating in that particular level of education. Gross enrollment rates compare all enrollments regardless of age (thus including under- and over-age students) to the "normal" age cohort; net enrollment rates compare only the enrollments from the normal cohort to the total population of that cohort.
8. The social and economic crisis reached its peak in December 2001, when President Fernando De la Rua was pressed to resign in a context of massive street demonstrations, supermarkets ransacking, and violent police repression. After De la Rua's resignation, the Congress appointed an interim President to complete the mandate and organize elections in the year 2003. The situation of crisis has resulted in increasing difficulties for national and provincial governments to invest in education, and to keep alive educational reform projects and compensatory programs.
9. Private schools play a significant role in Argentina. For example, in 1994 almost 30% of secondary students were enrolled in private institutions (Ministerio de Cultura y Educación, 1996a).
10. Different laws and decrees, dictated by the Congress and the executive power since the end of the 19th century, regulated the system before 1993.
11. The transfer of secondary schools to provinces was one of the measures that the national government had committed to in the negotiations for external credits with the International Monetary Fund between 1989 and 1991 (Senén González and Arango, 1997).
12. The modernization of provincial systems of educational administration was the focus of a specific program implemented from 1996 and partially funded by the World Bank (Decibe, 1998).
13. The Federal Council established 5 hours of instructional time per day and 180 days of class per year for the *Polimodal* cycle.
14. The content of the remaining 20% should be determined by each school according to its institutional project (see below the section on the new model of school management).
15. Under the De la Rua national administration (Dec. 1999-Dec. 2001), this deadline was extended to the year 2003.
16. According to Braslavsky (1998), "the Common Basic Contents are guided toward training in basic and fundamental skills, introducing many procedural contents, and placing emphasis on training in conceptual thinking more than on factual contents" (p. 308). Their design was the result of a process of consultation with experts from different fields of knowledge and with provincial MOEs.
17. In the Argentine system, most taxes are collected by the national government, and then distributed among the national and provincial administrations. As Senén González (1997) states, this system has been a source of conflicts and negotiation between the nation and the provinces. Since its creation in 1935, various legal instruments have changed the percentages for each level of government and the kinds of taxes included (Carciofi, et al, 1996). In addition, the national government has been allowed to make special contributions to provinces facing financial problems.
18. The national budget for education increased about30 % between 1993 and 1996 (Braslavsky, 1999). As Morduchowicz (1999) shows, this growth translated in an increasing transfer of funds from the national level to

provinces and schools in the areas of infrastructure, equipment, scholarships, development of school institutional projects, and teacher training.

19. However, in some cases, the 8th and 9th years have remained located in the secondary school buildings but under the direction of the EGB principal, which has resulted in problems of communication and coordination, and in teaching styles at those years that resemble more to those of the old secondary (Krichesky and Cappellari, 1999; Ministerio de Cultura y Educación, 2000b).
20. Even in the case of the province of Córdoba it can be questioned whether the change of structure represents the implementation of the EGB and the *Polimodal* or of a different model.
21. For the case of the Social Plan, Puiggrós (2000) states that clientelistic practices have influenced the way provinces and schools are selected.
22. For the province of Buenos Aires, for example, a World Bank document states that "progress in educational reform has been slower than hoped, due to Teacher and Staff Unions' resistance to change, long standing complaints over teacher salaries, and generally poor institutional capacities" (World Bank, 2000, p. 7).
23. According to Puiggrós (1999), the new national administration that took office in December 1999 had to accept the continuity of the Federal Law because many provincial governments (still in the hands of the Peronist party) were in favour of advancing in its implementation.
24. There have been questions about the quality and efficiency of the evaluation system (see *La Nación*, 11/23/98), and about its impact on schools and classrooms (Experton, 1999).
25. A World Bank (2000, p. 5) document, for example, states, "The ongoing reform of secondary education (in the province of Buenos Aires) has boosted the retention rates in the EGB3 cycle and is already showing a significant increase in enrollment for the Polimodal cycle."
26. By contrast, there is a general perception that from 2001 drop-out rates have been increasing in most provinces, particularly among teenagers, in the context of severe budgetary restrictions and the acceleration of social crisis. From 1998 to August 2002, the percentage of the population living in poverty increased from 27% to 53% according to official data (La Nación, 8/21/02).
27. Castro et al. hold that "In general, broad-based knowledge geared toward problem solving seems to be more valuable to graduates over their work lives than specific vocational skills" (p. 14). In their view, that is the approach employed in the *polimodales* of Argentina and Mexico.
28. Tedesco (1998) notes that the neglecting of teachers as active participants, a characteristic of many reforms in Latin America during the 1990s, needs to be reversed in order to improve the effectiveness of educational policies.

References

Acedo, C. and Belalcázar, C. (2001). "A Literature Review: Education Reform in the 1990s." The World Bank, Education Reform and Management Thematic Group.

Acedo, C. (2002) (ed.). *Case Studies in Secondary Education Reform*. Washington, DC: American Institutes for Research/USAID.

Aguerrondo, I., Xifra, S., and Delamer, L. (2000). *Validez nacional de Estudios y Certificaciones: Propuesta para su organización y gestión*. Document produced for the National MOE (Argentina). Buenos Aires.

Aguerrondo, I. (1996). *La escuela como organización inteligente*. Buenos Aires: Troquel.

Balduzzi, J. and Suárez, M. (1999). "La implementación del octavo año en las escuelas bonaerenses." Cuadernos de informes e investigaciones, SUTEBA.

Braslavsky, C. (2001). *Los procesos contemporáneos de cambios de la educación secundaria en América Latina: Análisis de casos en América del Sur*. IIEP.

Braslavsky, C. (1999, December). La reforma educativa en la Argentina: Avances y desafíos. *Propuesta Educativa*21, 80-88.

Braslavsky, C. (1998). Restructuring the Argentine educational system, 1984-1995. In Tulchin, J.S., with Garland, A.M. (Eds.), *Argentina: The challenges of modernization*. Wilmington, DE: Scholarly Resources Inc.

Braslavsky, C. and Tiramonti, G. (1990). *Conducción educativa y calidad de la enseñanza media*. Buenos Aires: Miño y Dávila Editores.

Carciofi, R. (coordinator), Cetrángolo, O., and Larrañaga, O. (1996). *Desafíos de la decentralización: Educación y salud en Argentina y Chile*. Santiago de Chile: CEPAL.

Carnoy, M. (1999) *Globalization and educational reform: What planners need to know*. Paris: UNESCO, Institute for International Planning.

Casassus, J. et al. (1998). *Estudio de seguimiento y evaluación de la Transformación Educativa en la Provincia de Buenos Aires*. Buenos Aires: Oficina Regional de UNESCO-Representación Argentina.

Castro, C. de M., Carnoy, M. and Wolff, L. (2000). *Secondary schools and the transition to work in Latin America and the Caribbean*. Washington, DC: IDB.

Cocorda, E. (1999) "La reforma educativa en las provincias argentinas. Retos metodológicos para un análisis en el marco de las relaciones intergubernamentales. Aspectos preliminaries." Paper presented at the IV Congreso Latinoamericano de Administración de la Educación, Universidad Nacional de San Martín, Buenos Aires, Argentina.

Consejo Federal de Cultura y Educación (1996). *La educación polimodal: Aportes para un Acuerdo Marco. Resolución Nro. 54/96*. Buenos Aires: Consejo Federal de Cultura y Educación.

Consejo Federal de Cultura y Educación (1994). *Pacto Federal Educativo*. San Juan, Argentina: Consejo Federal de Cultura y Educación.

CTERA. (2001). *Consulta nacional docente: La educación Argentina. Qué dicen los maestros?* [On-line]. Available: www.suteba.org.ar.

CTERA. (1997). *Situación de la educación en Argentina. Diagnóstico*. [On-line]. Available: http://www.wami.apc.org/ctera/diagno.html

Decibe, S. (1998). Transformar la educación para la calidad y la inclusión social: La experiencia argentina. Conference organized by the World Bank, *Construyendo el futuro de América Latina: Asociación público-privada para la educación*. Washington, DC.

Delich, A. (1999). Política educativa: Balance provisorio del fin de un ciclo. *Escenarios Alternativos*, 6, 60-69.

Dustchatzky, S. et al. (1999). *Aproximaciones al proceso de reforma en el Tercer Ciclo de la Provincia de La Pampa*. Buenos Aires: FLACSO.

Experton, W. (1999). *Desafíos para la nueva etapa de la reforma educativa en Argentina*. Washington, DC: The World Bank. LCSHD Paper Series No. 46.

Filmus, D. (1998). La descentralización educativa en el centro del debate. In *La Argentina que viene*. Buenos Aires: Norma.

FIEL/Centro de Estudios Públicos (2000). *Una educación para el siglo XXI: Propuesta de reforma*. Buenos Aires: FIEL.

Frigerio, G. (1995). *De aquí y de allá: Textos sobre la institución educativa y su dirección*. Buenos Aires: Kapelusz.

Fullan, M. (1993). *Change forces : Probing the depths of educational reform*. London and New York: Falmer Press.

García de Fanelli, A. M. (1997). Reformas en la educación básica y superior en la Argentina: Avances y restricciones en su implementación. In CEPAS (Ed.), *Los desafíos para el Estado en la Argentina actual*. Buenos Aires: Universidad de Buenos Aires.

Gozman, G. and Jacinto, C. (1999). *El desafío de la implementación de la escolaridad en el medio rural: el Programa del Tercer Ciclo de la EGB en escuelas rurales de Argentina*. Buenos Aires: Ministerio de Cultura y Educación/Instituto Internacional de Planeamiento Educativo-Buenos Aires.

Krichesky, M. and Cappellacci, I. (1999, December). Gestión curricular y riesgo pedagógico en EGB3: Un análisis de casos en Lomas de Zamora y La Matanza. *Propuesta Educativa* 21. 66-79.

La Nación (8/21/02). La crisis provocó que haya 5,2 millones de nuevos pobres. Buenos Aires.

La Nación (7/9/99). Los maestros percibirán un único pago de $360 en agosto. Buenos Aires.

La Nación (11/23/98). La pesada carga de las pruebas. Buenos Aires.

Levin, B. (2001).Conceptualizing the process of education reform from an international perspective. *Education Policy Analysis Archives*, 9 (14).

Llach, J. J., Montoya, S. and Roldán, F. (1999). *Educación para todos*. Córdoba: IERAL.

Ministerio de Educación (2000a). *El Tercer Ciclo desde la mirada docente: Avances y desafíos frente a la extensión de la obligatoriedad escolar*. Buenos Aires: Ministerio de Educación, Unidad de Investigaciones Educativas.

Ministerio de y Educación (2000b). *Implementación y Localización del Tercer Ciclo de EGB: Las prescripciones y su impacto en los actores institucionales*. Buenos Aires: Ministerio de Educación, Unidad de Investigaciones Educativas.

Ministerio de Educación (2000c). *La estructura curricular básica del Tercer Ciclo de la EGB en ocho jurisdicciones*. Ministerio de Educación, Unidad de Investigaciones Educativas.

Ministerio de Educación (2000d). *Estado de situación curricular de las provincias*. Buenos Aires: Subsecretaría de Educación Básica.

Ministerio de Educación (2000e). *El trabajo docente en el Tercer Ciclo de la EGB. La normativa y las instituciones*. Ministerio de Educación, Unidad de Investigaciones Educativas.

Ministerio de Cultura y Educación (1997). *Argentine education in the society of knowledge*. Presented at the 29th Session of the General Conference UNESCO, Paris, October 1997.

Ministerio de Cultura y Educación (1996a). *Censo nacional de docentes y establecimientos educativos '94: Matrícula escolar*. Buenos Aires: Ministerio de Cultura y Educación.

Ministerio de Cultura y Educación (1996b). *Condiciones básicas institucionales: Nuevos contenidos en una escuela diferente*. Buenos Aires: Ministerio de Cultura y Educación.

Ministerio de Cultura y Educación (1996c). *Aspectos estructurantes de la organización y la gestión de la escuela: el tiempo, el espacio y los agrupamientos*. Buenos Aires: Ministerio de Cultura y Educación.

Minujin, A. and Kessler, G. (1995). *La nueva pobreza en la Argentina*. Buenos Aires: Editorial Planeta.

Morduchowicz, A. (1999). *Una mirada desde los costos del sistema educativo argentino*. Informe Presentado a la Comisión de Ecología y Desarrollo Humano del Honorable Senado de la Nación. Buenos Aires.

Puiggrós, A. (2000). Educación básica y media en la Argentina de comienzos del siglo XXI. *Revista Aportes*, 15.

Puiggrós, A. (1997). *La otra reforma: Desde la educación menemista hasta el fin de siglo*. Buenos Aires: Galerna.

Roggi, L. (2001). Esta Resolución no ayudará a mejorar la situación. *Novedades Educativas*,13 (123). Buenos Aires.

Senén González, S. (2000). Argentina: Actores e instrumentos de la reforma educativa. Propuestas del centro y respuestas de la periferia. *Revista Alternativas* (Universidad Nacional de San Luis).

Senén González, S. and Arango, A. (1997). La descentralización educativa: Política educativa o política fiscal? In Oszlak, O. (Ed.), *Estado y sociedad: Las nuevas reglas del juego*, 1. Buenos Aires: CBC-CEA.

Senén González, S., with the collaboration of M.A. Sendón (1997). *Ajuste y reforma educativa: dos lógicas en pugna*. III Congreso Latinoamericano de Administración de la Educación, UNICAMP, Campinas.

Stubrin, A. (1999). Educación, igualdad, solidaridad y cohesión. *Escenarios Altenativos*, 6, 30-41.

SUTEBA (Sindicato Unificado de los Trabajadores de la Educación de la Provincia de Buenos Aires) (1999). *Una reforma que no Transforma*. [On-line]. Available: www.suteba.org.ar. Extracted in February, 2001.

Tedesco (1998). Desafíos de las reformas educativas en América Latina. *Propuesta Educativa*,19, 19-23. Buenos Aires.

Tiramonti, G. and Braslavsky, C. (1995). Quiénes ofrecen educación en la Argentina de hoy. In G. Tiramonti, C. Braslavsky and D. Filmus (Eds). *Las transformaciones de la educación en diez años de democracia*. Buenos Aires: Tesis Norma.

Tiramonti, G. (1996). *Los nuevos modelos de gestion educativa y su incidencia sobre la calidad de la educacion*. Buenos Aires: Flacso/Serie documentos e Informes de Investigación, 211.

Tiramonti, G. (1995). Quiénes van a la escuela hoy en Argentina? In G. Tiramonti, C. Braslavsky and D. Filmus (Eds.). *Las transformaciones de la educación en diez años de democracia*. Buenos Aires: Tesis Norma.

Wolff, L., and Castro, C. de M. (2000). *Secondary education in Latin America and the Caribbean: The challenge of growth and reform*. Washington, DC: IDB.

World Bank. (2000). *Argentina, AR BA Secondary Education II: Project concept document.* Latin America and Caribbean Region LCSHE.

World Bank. (1995). *Staff appraisal report. Argentina: Decentralization and improvement of secondary education and polymodal education development project.* Human Resources Operation Division. Report No. 14559-AR.

World Bank. (1994). *Staff appraisal report. Argentina: Decentralization and improvement of secondary education project.* Human Resources Operation Division. Report No. 12993-AR.

About the Authors

Jorge M. Gorostiaga
University of Pittsburgh
E-mail: gorosti@pitt.edu

Jorge M. Gorostiaga is a doctoral candidate in the Social and Comparative Analysis in Education program at the University of Pittsburgh. He has taught at universities in Argentina. His research focuses on educational reform. He is the co-editor (with Mark B. Ginsburg) of Dialogue among researchers, policy makers and practitioners: International perspectives on the field of education (Routledge Falmer, 2003).

Clementina Acedo
5K34 Posvar Hall
Institute for International Studies in Education
University of Pittsburgh, PA 15260
E-mail: cacedo@pitt.edu

Clementina Acedo is Assistant Professor in Administrative and Policy Studies in the School of Education, University of Pittsburgh. She has worked as Education Specialist for the World Bank and as Researcher, Consultant and Professor in Caracas, Venezuela. Her recent research includes a book and articles on: Secondary education reform in the nineties in Argentina, Egypt, Kenya, Romania and Indonesia and topics of justice and development in social and educational policy.

Susana E. Xifra
Buenos Aires, Argentina
E-mail: xijen@radar.com.ar

Susana E. Xifra is a consultant at the International Institute of Educational Planning-UNESCO (Buenos Aires). She has taught at primary, secondary and university institutions. She has worked as an educational official in Argentina at provincial and national levels.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
UmeÅ University

Thomas Mauhs-Pugh
Green Mountain College

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | book reviews | subscribe | search

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 1434 times since June 12, 2003

Volume 11 Number 18 **June 12, 2003** **ISSN 1068-2341**

An Examination of the Longitudinal Effect of the Washington Assessment of Student Learning (WASL) on Student Achievement

Donald C. Orlich
Washington State University

Citation: Orlich, Donald C. (2003, June 12). An examination of the longitudinal effect of the Washington Assessment of Student Learning (WASL) on student achievement. *Education Policy Analysis Archives, 11*(18). Retrieved [date] from http://epaa.asu.edu/epaa/v11n18/.

Abstract

Linn, Baker and Betebenner (2002) suggested using the effect size statistic as a measure of adequate yearly progress target (AYPT) as is required by PL 107-110. This paper analyzes a four-year data set from the required high-stakes test--Washington Assessment of Student Learning—using effect size as the AYPT metric. Mean scale scores for 4th, 7th and 10th grade reading and mathematics were examined. Nominal descriptors suggested by Cohen (1988) were applied and showed no yearly effect in student achievement as a function of the WASL. Comparing the 1998 scale scores to those of 2001 showed a small effect. However, manipulating the effect size criterion from 0.20 to 0.05 did show small yearly effects in student achievement. Meeting AYPT objectives will be a problem of defining the standard as yearly score fluctuations occur. The educational research community should challenge the statistical logic associated with setting AYPT's.

A statistical and accountability dilemma has emerged due to the passage of the "No Child Left Behind Act of 2001" (PL 107-110). States are now forced by federal law to show adequate student yearly progress targets, which will be met through high-stakes testing. Several states have constructed their own accountability systems that feature criterion-referenced assessments. However, student test scores tend to display characteristics of norm-referenced tests, i. e., normal distributions. The use of effect size statistics, which is herein applied to one state, has been suggested as one means of determining the required adequate yearly progress targets (Linn, Baker & Betebenner, 2002).

The Washington State Model (WASL)

The State of Washington established the Washington Assessment of Student Learning (WASL) as its accountability tool. The WASL is keyed to the state's standards called "Essential Academic Learning Requirements." The WASL is used to test all 4th, 7th and 10th graders in mathematics, reading, and writing. The 5th, 8th and 10th graders will be assessed in science. Listening is also assessed. Using the data collected from the 1998 through 2001 WASL administrations; the writer calculated effect sizes (see Cohen, 1988) to observe trends.

The purpose of this study is to determine the effect on student achievement as a consequence of the longitudinal administration of the Washington Assessment of Student Learning (WASL) the state mandated high-stakes test. The WASL scale score means and standard deviations were available for the years 1998, 1999, 2000 and 2001 for mathematics and reading and are shown in Table 1. The average number of students taking the WASL math and reading tests during the four year period for grades 4, 7 and 10, respectively, are

70,431; 72,864 and 66,856. These three combined totals account for 21 percent of the state's total 2001-02 K-12 student population of 1,010,424 (Education Profile, 2002). The number of WASL test-takers is significant.

Table 1
Yearly Means and Standard Deviations for 4th, 7th, and 10th Grade Mathematics and Reading Scale Scores On the Washington Assessment of Student Learning (1998-2001)

Grade Level and Subject	Spring 1998 Means	Spring 1998 Standard Deviations	Spring 1999 Means	Spring 1999 Standard Deviations
4 - Mathematics	383.5	32.2	386.5	33.9
4 - Reading	402.1	19.3	404.2	19.5
7 - Math	357.4	46.4	364.7	52.0
7 - Reading	390.1	20.1	393.1	20.2
10 - Math	N/R	N/R	382.2	42.8
10 - Reading	N/R	N/R	402.8	29.5

Grade Level and Subject	Spring 2000 Means	Spring 2000 Standard Deviations	Spring 2001 Means	Spring 2001 Standard Deviations
4 - Mathematics	391.2	34.9	393.3	34.9
4 - Reading	407.3	19.6	405.7	18.6
7 - Math	369.1	53.6	368.7	51.6
7 - Reading	393.8	20.9	394.5	20.6
10 - Math	387.6	40.0	390.8	41.1
10 - Reading	407.3	30.2	410.0	30.5

All means and standard deviations are from files of Office of State Superintendent of Public Instruction, Olympia, Washington.

An inspection of the scale score means shows a rather small incremental increase in most means. However, there is a scale point decline of 0.4 in the mean of grade 7, 2001 math scores compared to 2000. A similar decline is noted in 2001 for grade 4 reading, where the mean scale score dropped by 1.6 points compared to 2000.

These patterns have been praised by state policy makers as showing evidence of student progress. However, are the scores truly reflective of student achievement? To answer that question, I use the statistical test of *effect size*.

Effect size

The *effect size* is a method by which to judge the relative learning worth from independent samples (see Bloom, 1984; Cohen, 1988; Glass, 1980; Marzano et al., 2001; & Walberg, 1999). In this case, what evidence is there that administering and teaching to the WASL has a positive impact on student achievement?

Cohen (1988) defined an effect size as the difference between two means divided by the standard deviation of either group. With independent samples, such as the WASL, one can determine the effect sizes by comparing the means of two different years. Cohen also suggested that the relative efficacy of an effect could be stated in nominal terms. If an effect size (ES) were at least a 0.2, it was labeled as *small*. An ES of at least 0.5 was labeled as *medium;* while and ES of 0.8 or greater was *large*. Thus, an effect size of 0.2 is required to show efficacy of learning. Table 2 shows the effect size calculations and nominal descriptors for this study. In all calculations a uniform method was used. The earlier year is the control, while the latter year of the pair is the experimental group. Standard deviations are from the control years.

Table 2
Effect Size Calculations for 4th, 7th, and 10th Grade Mathematics and Reading Scores On the Washington Assessment of Student Learning (1998-2001)

Grade Level And Subject	1999/1998 Effect Size	Effect	2000/1999 Effect Size	Effect	2001/2000 Effect Size	Effect
4 - Math	0.09	None	0.14	None	0.06	None
4 - Reading	0.11	None	0.16	None	-0.08	Negative
7 - Math	0.16	None	0.08	None	-0.01	Negative
7 - Reading	0.15	None	0.03	None	0.05	None
10 - Math	N/R	N/R	0.13	None	0.08	None
10 - Reading	N/R	N/R	0.15	None	0.09	None

The effect is described in nominal terms as per Jacob Cohen's (1988) definitions.

Discussion of Data Sets

Table 2 shows the effect sizes for the 4th, 7th and 10th grade mathematics and reading scores yearly from 1998 to 2001. Examining Table 2, you may note that at the 4th grade level, five scores show *no effect* in achievement, while there is *one negative learning effect* on grade 4 reading in 2001, that is, a decline in achievement.

The grade 7 pattern is similar showing *no effect* on five of the six scores and one *negative effect* in mathematics for 2001. The grade 10 results show *no effect* on mathematics and reading scores in all cases.

If one were to use an effect size of 0.05, which would account for a two percentile gain, as the Adequate Yearly Progress Target (AYPT), suggested by Linn, Baker, and Betebenner (2002), then 13 of the 16 scores would meet that target. However, using Cohen's (1988) definitions the 16 scores would show no effect and not meet the target. Setting the criterion measure of an adequate yearly progress target (AYPT) may become a major problem of definition. This situation may further complicate the implementation of AYPT policy. It appears that further analysis of setting AYPT's and field-studies are essential.

The average percentile gain during the four-year period of this study (1998-2001) for grades 4, 7 and 10 in math and reading was 3.3. That number would correspond to an effect size of 0.08. That effect size would exceed the AYPT suggested by Linn et al. (2002), but not the nominal ES suggested by Cohen. These differences will be explored further.

Now examine Tables 3 and 4, which display the data and effect sizes by comparing the 1998 WASL scores for grades 4 and 7 to those of 2001; and grade 10 for 1999 and 2001. Using Cohen's classifications, five of the six comparisons show a small effect and only one with no effect. It appears to take several years to show any effect. Using WASL math results released in the SRI International report (2002) showed that for grade 4 from 1997-2002, the ES was 0.79. The ES for grade 7 from 1998-2002 was 0.31; while the ES for grade 10, 1999-2002, was 0.14. These data are inconclusive, but do suggest that like national norm-referenced tests, state criterion-referenced tests may take several years to show a positive impact on student achievement.

Camara and Powers (1999) concluded that coaching students for the SAT does in fact increase a student's SAT score. Washington State teachers have had at least four years experience in preparing for and teaching to the WASL. Further, between June 21 and June 30, 2001, the State Superintendent of Public Instruction selected 175 classroom teachers to attend a special, all expenses paid, WASL assessment training program in Mesa, Arizona (Brown, 2001, May 17). The impact of that training and student familiarity with the WASL appear to be similar to the SAT coaching findings reported by Camara and Powers (1999), especially at grade 4 where teachers have had over six years of "practice." Additionally, the state superintendent of public instruction initiated a "School Improvement Specialist" program in 2001-02. About 200 selected individuals are being paid $30,000 a school year to work 1.5 days per week (or up to $90,000 for 4.5 days) to help teachers and schools "improve student performance." No independent evaluation of this multi-million dollar expenditure has yet been conducted.

Table 3
Means, Standard Deviations, and Effect Sizes for 4th, and 7th Grade Mathematics and Reading Scale Scores on the Washington Assessment of Student Learning Comparing (1998—2001)

Grade Level and Subject	Spring 1998 Means	Spring 1998 Standard	Spring 2001 Means	Spring 2001	Effect Size	Effect

		Deviations		Standard Deviations		
4 - Mathematics	383.5	32.2	393.3	34.9	0.30	Small
4 - Reading	402.1	19.3	405.7	18.6	0.19	None
7 - Math	357.4	46.4	368.7	51.6	0.24	Small
7 - Reading	390.1	20.1	394.5	20.6	0.22	Small

Table 4
Means, Standard Deviations, and Effect Sizes for 10th Grade Mathematics and Reading Scores on the Washington Assessment of Student Learning Comparing (1999—2001)

Grade Level and Subject	Spring 1999 Means	Spring 1999 Standard Deviations	Spring 2001 Means	Spring 2001 Standard Deviations	Effect Size	Effect
10 - Math	382.2	42.8	390.8	41.1	0.20	Small
10 - Reading	402.8	29.5	410.0	30.5	0.24	Small

It appears that nearly four years would be required to show a small effect on student test performance; thus AYPTs may be elusive and troublesome. And how does a state account for a negative effect size year as the Washington data illustrate? The Washington State reform movement was legislated in existence in 1993. Thus, one could argue that during an eight year period little impact on student achievement is shown for the estimated $1 Billion cost for the state's total school reform package.

A question needs resolution: "Is the small effect shown on the four year comparisons a function of increased teacher knowledge and possible compromise of WASL questions, or is the effect a function of growth in student academic achievement?" Obviously that question is not answered in this paper, but must be explored.

Further Implications

This article only examines the effect size of student achievement as a consequence of longitudinal high-stakes testing in one state. To illustrate the gravity of the problem let us also examine the percent of students meeting the arbitrarily set standard scale score in math and reading for grades 4, 7 and 10 from 1998-2002. (See Table 5.)

Table 5
Percent of Students Meeting and Not Meeting Standard for 4th, 7th, and 10th Grade Mathematics and Reading on the Washington Assessment of Student Learning (1998—2002)

Grade Level And Subject	1998-99 % Meeting Standard	1999-2000 % Meeting Standard	2000-2001 % Meeting Standard	2001-2002 % Meeting Standard
4 - Math	37.3	41.8	43.4	51.8
4 - Reading	59.1	65.8	66.1	65.6
7 - Math	24.2	28.2	27.4	30.4
7 - Reading	40.8	41.5	39.8	44.5
10 - Math	33.0	35.0	38.9	37.3
10 - Reading	51.4	59.8	62.4	59.2

Source: State Superintendent of Public Instruction, Education Profiles, Olympia, WA, 1998-2002.

The range of percentages of students meeting the standard for math in grade 4 is from 37.3 in 1998-99 to 51.8 percent in 2001-02. For grade 7, the range is 24.2 to 30.4 percent. And in grade 10, the range in math is from

33.0 to 37.3 percent. Considering that yet about one-half of the fourth graders and almost two-thirds of the seventh and tenth graders do *not* meet the standard (fail) policy makers should be alarmed. In deed, on November 30, 2002 writers of the SRI International report concluded "The analyses further suggest that the grade 7 test was more challenging for the 7th graders than the grade 10 test was for the 10th graders" (page ii). The report also noted that several WASL test items on the 4th grade math test were not aligned with the Essential Academic Learning Requirements.

The state superintendent of public instruction informed the standard-setting groups that the WASL cut scores and standard-setting are guided by the belief that, "In all content areas the standard should reflect what a well taught, hardworking student should know and be able to do near the end of grade [4,7, or 10]" (SRI, 2002, p. 20). It is obvious that developmental, cognitive or behavioral perspectives are not being reflected in that guiding principle. Is faculty psychology now in vogue?

Orlich (2000) analyzed the 4th grade practice WASL items by using the developmental scales published by Epstein (2002, p. 184), Bloom's Taxonomy (1954), and the NAEP scales (Campbell et al, 1998) and concluded that the bulk of items on the 4th grade WASL were well beyond their developmental level. Results of the 1998-99 WASL results confirmed that conclusion. Thus, the very high percentages of students "not meeting the WASL standard" may be traced to developmentally inappropriate items.

All children must take the WASL, with very few exceptions. The math effect size differences between grade-four minority children and white/Caucasian children range from 0.75 to 0.80, or about 28 percentile. For Special education the effect size difference between that population and white/Caucasian children is 0.96 or a 33 percentile difference (see Taylor, 2001, Tables 8-10 and 8-12). The latter is nearly one full standard deviation on the WASL. Finally, there is a correlation between reading and math WASL scores of 0.73 (Abbott and Joireman, 2001). Is math or reading being tested?

The cost-effectiveness of the WASL may be determined by the actual contract costs of the WASL with the Riverside Publishing Company. The initial contract was for about $40 Million. The renewal (2001-05) is $61,673,910 (Contract no.120-761, 2001). Thus the WASL cost is about $102 Million per se. That figure does not include teacher salaries for time spent to prepare students or to administer the WASL, plus other bureaucratic costs associated for its administration. With an average 3.3 percentile gain per year, the cost per one percentile gain is about $11 Million per year. Obviously, the cost-benefit calculation is challengeable; but no matter because the cost to meet any AYPT, as is mandated by federal laws (PL107-110), will be staggering. More importantly the money goes only to the test publishers. Not one dollar of the WASL reform expenditure goes to teachers to aid their instructional efforts.

Are the WASL Scores aberrations? Apparently not: Linn and Haug (2002) examined Colorado school buildings test scores over a four-year period and concluded the following:

"The performance of successive cohorts of students is used in a substantial number of states to estimate the improvement of schools for purposes of accountability. The estimates of improvement, however, are quite volatile. This volatility results in some schools being recognized as outstanding and other schools identified as in need of improvement simply as the result of random fluctuations. It also means that strategies of looking to schools that show large gains for clues of what other schools should do to improve student achievement will have little chance of identifying those practices that are most effective. On the other hand, schools that are identified as 'in need of improvement' will generally show increases in scores the year after they are identified simply because of the noise in the estimates of improvement and not because of the effectiveness of the special assistance provided to the schools or pressure that is put on them to improve." (p. 35).

Similarly, Darling-Hammond (2003) reported that a doubt must be cast on state test gain scores because in Texas, students showed gains on the state mandated assessment, but did not make comparable gains on national standardized tests or the Texas college entrance test.

Conclusion

Using an effect size measurement and Cohen's (1988) nominal definitions, there is no effect, that is, no positive impact on yearly student achievement as a consequence of the longitudinal administration of the Washington Assessment of Student Learning (WASL). However, over a four-year period a small effect size does emerge. The results of this study parallel the findings of Amrein and Berliner (2002a) who analyzed the consequences of 18 states with high-stakes tests. They reported that in 17 of the 18 states, student learning remained at the same level as it was before the policy of high-stakes tests was instituted.

In two separate studies, the first of 28 states with high-stakes tests, Amrein and Beliner (2002b) concluded that these tests do little to improve student achievement. In a second study of 17 states (2002c) they concluded that high-stakes tests may actually worsen academic performance and exacerbate dropout rates. The affective dimensions of high-stakes tests should be of great concern to policy makers and educators alike.

Washington State policy makers must re-examine the intent of the WASL and the empirical data sets that analyze it to determine its educational worthiness and continued fiscal support (Orlich, 2000; Abbott & Joireman, 2001; Basarab, 2001; Fouts, 2002; & Keim, 2002). At the federal level there is need to examine the practicality, reasonableness and statistical logic of setting adequate yearly progress targets. The experience in the state of Washington apparently shows that setting AYPT's may not only be an assessment fallacy, but a gross misapplication of adapting the banking practice of applying compound interest calculations to human cognition. Is educational reform anything you can get away with?

Note

The Author expresses appreciation to colleagues at Washington State University, V. S. Manaranjan and Michael S. Trevisan for their critique, insights, and suggestions relating to this paper; and to the anonymous referees for their constructive feedback.

References

Abbott, M. L. & Joireman, J. (2001, July). The relationships among achievement, low income, and ethnicity across six groups of Washington state students. Lynnwood, WA: Washington School Research Center, Technical Report #1.

Amrein, A. L. & Berliner, D. C. (2002a, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved April 10, 2002 from http://epasa.asu.edu/epaa/v10n18/

Amrein, A. L. & Berliner, D. C. (2002b, December). The impact of high-stakes tests on student academic performance: an analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP test results in states with high school graduation exams. Education Policy Research Unit, Education Policy Studies Laboratory, Arizona State University, Tempe, AZ. (EPSL-0211-126-EPRU).

Amrein, A. L. & Beliner, D. C. (2002c, December). An analysis of some unintended and negative consequences of high-stakes testing. Education Policy Research Unit, Education Policy Studies Laboratory, Arizona State University, Tempe, AZ. (EPSL-0211-125- EPRU).

Basarab, S. (2001, February). An overview of student assessment in Washington state. Unpublished Report. Citizens United for Responsible Education (CURE). Burien: WA. Web site of CURE is at: **http://www.eskimo.com/~cure/**

Bloom, B. S. et al. (1956) *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13* (6), 4-16.

Brown, D. (2001, May 17). Staffing paper: MEA for Mesa, Arizona trip. Olympia, WA: Office of State Superintendent of Public Instruction files.

Camara, W. J. & Powers, D. (1999, July). Coaching and the SAT. Society for Industrial Psychology Inc. Retrieved August 27, 2002, from http://stop.org/tip/backissues/tipjuly99/4camara.htm.

Campbell, J. R., Voelke, K. E., & Donahue, P. L. (1998, August). *Report in brief: NAEP trends in academic progress, revised*. Washington, DC: National Center for Education Statistics, (NCES no. 998-530).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd Ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Contract No. 120-761. (2001, September 20). Riverside Publishing Contract with Superintendent of Public Instruction for WASL implementation.

Darling-Hammond, L. (2003). Standards and assessments: Where we are and what we need. *Teachers College Record*. Retrieved February 27, 2003 from http://www.tcrecord.org/Contentasp?ContentID=11109

Education Profile. State Superintendent of Public Instruction. Olympia, WA. Retrieved at http://www.k12.wa.us/edprofile/state

Epstein, H. T. (2002). Biopsychologicalaspects of memory and education. In S. P. Shohov, (Ed.). *Advances in Psychology Research*, vol. 11. New York: Nova, 181-186.

Fouts, J. T. (2002, April). The power of early success: A longitudinal study of student performance on the Washington assessment of student learning, 1998-2001. Lynnwood, WA: Washington School Research Center, Research Report #1.

Glass, G. V. (1980). Summarizing effect sizes. In R. Rosenthal, (Ed.). *New directions for methodology of social and behavioral science: Quantitative assessment of research domain.* San Francisco: Jossey-Bass.

Keim, W. G. (2002). School accountability and fairness: A policy study of the 2000 Washington state school accountability criteria. Pullman, WA: Washington State University, Unpublished Doctoral Dissertation.

Linn, R. L., Baker, E. V., & Betebenner, D. W. Accountability systems: Implications of requirements of the no child left behind act of 2001. *Educational Researcher*, 31 (6), 3-16.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

Orlich, D. C. (2000). A critical analysis of the grade four Washington assessment of student learning. *Curriculum In Context, 27* (2), 10-14

No Child Left Behind Act of 2001, Public Law No. 107-110, 115 Stat. 1425 (2002).

SRI International. (2002, November 30). A review of the Washington assessment of student learning in mathematics: Grades 7 and 10. Final Report. SRI Project Number: P12089. Olympia, WA: The Office of Superintendent of Public Instruction.

Taylor, C. S. & Lee, Y. (2001, October 18). Washington assessment of student learning grade 4, 2000 technical report. Olympia, WA: Office of the State Superintendent of Public Instruction.

Walberg, H. J. (1999). Productive teaching. In H. C. Waxman and H. J. Walberg, (Eds.), *New directions for teaching practice and research* (pp. 75-104). Berkeley: CA: McCutchan Publishing Corporation.

About the Author

Donald C. Orlich
Science Mathematics Engineering Education Center
Washington State University
Pullman, Washington 99164-4237
Telephone: (509) 335-4844 FAX: (509) 335-7389

E-mail: dorlich@wsu.edu

Donald C. Orlich is Professor Emeritus, Science Mathematics Engineering Education Center at Washington State University, Pullman. His academic focus is science education and instruction. He is senior co-author with R. Harder, R. Callahan, M, Trevisan and A. Brown of the forthcoming 7th edition of *Teaching Strategies: A Guide to Effective Instruction,* Houghton Mifflin, 2004. He authored *Designing Successful Grant Proposals,* ASCD 1996/2002. On March 2001, the 160,000 member ASCD honored him with the "The Affiliate Publications Award for Outstanding Affiliate Article" for "A Critical Analysis of the Grade Four Washington Assessment of Student Learning."

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 1057 times since June 30, 2003

Volume 11 Number 19 **June 30, 2003** **ISSN 1068-2341**

Using Large-Scale Research to Gauge the Impact of Instructional Practices on Student Reading Comprehension: An Exploratory Study

Harold Wenglinsky

Baruch College

Citation: Wenglinsky, H.. (June 30, 2003).Using Large-Scale Research to Gauge the Impact of Instructional Practices on Student Reading Comprehension: An Exploratory Study. *Education Policy Analysis Archives, 11*(19). Retrieved [date] from http://epaa.asu.edu/epaa/v11n19/.

Abstract

Small-scale research has identified classroom practices that are associated with high student performance in reading comprehension. It is not known, however, whether these findings generalize to larger samples and populations, as most large-scale studies of the impact of teaching on student performance do not include measures of classroom practices. Generalizing to larger populations is particulary important at a time when policies national in their scope are calling for "scientifically-based" instruction in reading. The current study explores the possibility of using large-scale data and methods to study classroom practices in reading comprehension. It finds that such studies are both feasible and necessary. They are feasible insofar as it proved possible to collect and analyze data on classroom practices and student reading comprehension, and discern substantial effects of the one on the other. They are necessary insofar as the study confirmed the effectiveness of some classroom practices but not others. It therefore cannot be assumed that the findings of small-scale studies generalize to large populations.

Introduction

With the passage of the Federal Reading First program and parallel state efforts to improve student reading skills, policy makers and educators have been looking for "scientifically-based" materials, professional development and instruction. While the research base on early reading skills such as phonics has been found to be substantial, there is a lack of conclusive evidence on effective instruction in more advanced skills such as reading comprehension.

This research gap is, for the most part, attributable to the small-scale nature of research that seeks to identify effective techniques for teaching reading comprehension. Over the last forty years, researchers have conducted a host of small-scale studies identifying certain classroom practices as effective in teaching reading comprehension. The findings of these studies have been remarkably consistent, with the same set of classroom practices again and again appearing to be related to student reading comprehension performance. The strength of these studies lies in their high level of internal validity. Many use experimental designs, and many others are quasi-experimental. These studies also possess a common shortcoming, however. Because the development of a robust design is extremely labor intensive, such studies tend to be small in scale, limited to a few classrooms or schools. The degree to which these studies apply to large populations is therefore not known.

Large-scale research does not provide much information on classroom effects. The common method for gauging the impact of teaching on student performance using large-scale data is known as the production function. It involves collecting observational data on large numbers of teachers and students and then using the technique of regression analysis, which relates teacher characteristics to student performance. Most production function studies do not measure classroom practices, due to the difficulty of measuring them for large numbers of teachers. And most have not found a clear and consistent relationship to student outcomes for the teacher characteristics they do measure.

The present study addresses methodological problems common in the production function literature to demonstrate the possibility of using large-scale data to study classroom practices. The problems addressed here include the lack of measurement models, the low validity and reliability of teacher self-reports and the interdependence of many independent variables. This is accomplished by using national data on 7,194 fourth graders and their teachers from the 2000 National Assessment of Educational Progress (NAEP). The study relates teachers' classroom practices, as well as their background characteristics, to student performance on a reading comprehension assessment, taking into account student background characteristics.

The study finds that the addition of classroom practices to large-scale models of reading performance is vital to the successful isolation of teacher effects. Once such variables are introduced, teacher effects can prove quite substantial, nearly as large as student background effects. The study also finds that testing small-scale results with large-scale data is crucial for establishing the effectiveness of specific classroom practices. A link was confirmed between some but not all classroom practices and student performance. It would be premature, however, to draw substantiative conclusions about effective reading practices from these findings. Rather, the purpose of this paper is to draw methodological conclusions about the viability of large-scale research on instructional practice. Before addressing these conclusions, however, it is worth reviewing key findings from the prior literature and describing in some detail how the current study was conducted.

Background

Intensive study has been devoted to classroom effects on reading. The reports of the National Research Council (Snow et al., 1998) and the National Reading Panel (2000) have identified hundreds of studies of how classroom practices affect reading performance. Much of this research, however, focuses on the early stages of reading skill acquisition, such as phonemic awareness and word recognition, with much less on reading comprehension (Snow et al., 2000). Nonetheless, the body of research on reading comprehension is substantial enough to make it possible to identify seven kinds of practices that are consistently associated with improvements in student reading comprehension.

First, students perform better when explicitly taught metacognitive skills. Metacognitive skills are the ways in which readers glean meaning from texts. In a series of studies, Durkin (1978, 1981) found that teachers, in explicating texts, rarely instruct students in methods of explication. In the wake of this finding, a variety of approaches to teaching metacognitive skills were developed, including reciprocal teaching, questioning and direct instruction. Research on these techniques has generally found positive effects on reading comprehension (Cross & Paris, 1988; Rosenshine & Meister, 1994; Hansen & Pearson, 1983; Wharton McDonald et al., 1998; Kaniel et al., 2000; Mueller, 1997; Alfassi, 1998).

Second, students seem to perform better when reading and writing instruction are integrated. After years of these skills being taught separately, educators called for their integration in the 1980's. Later, reading and writing were combined under the single heading of language arts, as reflected in many of the recently promulgated state academic standards. Research on using writing to improve reading comprehension has generally supported the approach (Wharton-McDonald et al., 1998; Cantrell, 1999; Knapp et al., 1995).

Third, research on the texts students read has documented advantages to using trade books rather than basal readers. Basal readers tend to abridge texts to maximize accessibility, whereas trade books are real world texts, and are often chosen to convey content rather than just reading skills. The use of trade books has been found to increase student motivation as well as to improve reading comprehension skills (Popplewell & Doty, 2001; Guthrie, 2001; Guthrie et al., 2000; Guthrie, 2000; Guthrie et al., 1999; Guthrie, 1998; Guthrie 1996).

Fourth, students seem to benefit from time spent reading in class. Research on time on task has suggested that more time spent teaching reading is associated with improved performance. Particularly strong effects have been found for time spent in the act of reading, with oral reading appearing more beneficial than silent reading (Topping & Paul, 1999; National Reading Panel, 2000).

Three other instructional techniques supported by the literature are having students work in groups, involving parents and using authentic assessments to measure student progress. Supposedly, students learn more from one another and are more motivated when engaging in group work. A variety of parental involvement activities, from checking homework to reading together, have been found to be conducive to improved reading performance (Epstein, 2001; Epstein & Dauber, 1994). And measuring student performance through tasks that are as similar as possible to class work and homework seems to be more effective than having students take traditional multiple-choice or short-answer tests.

While small-scale research has succeeded in identifying numerous ways in which teachers affect student performance, large-scale research has generally not been able to confirm these findings. Beginning with the Equality of Educational Opportunity Study (Coleman et al., 1966), production function studies have found that most teacher effects are overwhelmed by student effects. Of the hundreds of production functions estimated in the wake of the Coleman Report, less than one-third could discover a link between student outcomes and teacher experience, less than one-quarter could do so for teachers salaries, and just one in ten could do so for educational attainment. Two types of teacher effect did prove more robust. Many studies have isolated modest effects for teachers' majoring in the subject they teach and teacher scores on basic skills tests. The relative lack of large-scale studies confirming teacher effects, however, has to led to meta-analysis of them coming to divergent conclusions, some accepting and others questioning the existence of teacher effects (Hanushek, 1997; Hanushek, 1996a; Hanushek, 1996b; Hanushek, 1989; Greenwald, Hedges & Laine, 1996; Hedges & Greenwald, 1996; Hedges, Greenwald & Laine, 1994).

The disappointing results of large-scale studies may stem from their various methodological shortcomings. First, these studies tend to focus on teacher effects that are relatively easy to measure with large-scale data, namely teacher background characteristics such as education level or college major. Such studies thus tend not to measure the effects that small-scale research has found to be substantial. Second, such studies tend to lack measurement models; they assume variables are perfectly measured and do not develop constructs from multiple indicators. Yet measurement error in teacher self-reports of behavior and background is substantial, and can be minimized through multiple indicators (Mayer, 1999). Third, such studies tend not to relate independent variables to one another, when small-scale research suggests that teacher variables are very much affected by student background and school context.

A few large-scale studies do relate classroom practice to student performance in mathematics and science, and reveal substantial classroom effects. The nationally representative National Educational Longitudinal Study (National Center for Education Statistics, 1996) found that an emphasis by teachers on conveying higher-order thinking skills was positively associated with student performance in math but not in science. A study representative of the state of California (Cohen & Hill, 2000) found that reform-minded classroom practices were positively associated with student mathematics performance. And an analysis of the nationally representative 1996 National Assessment of Educational Progress in Mathematics (Wenglinsky, 2001) found that an emphasis on conveying higher-order thinking skills, engaging in hands-on learning activities, and receiving professional development to address special populations of students were all positively related to math scores. The fact that all three of these studies uncovered substantial teacher effects when classroom variables were included suggests the need for similar work in the area of reading. It is to this work that we now turn.

Research Questions, Data and Method

The exploratory study described here is designed to address two methodological research questions suggested by the prior literature. First, do the classroom practices identified as important by the small-scale literature prove to be uniformity related to student reading performance? If all are confirmed, large-scale research can be said to add little to what is already known. If it proves, however, that some practices are confirmed while others are not, this finding would suggest the need to conduct an independent program of studies using large-scale data. Second, does the addition of classroom practices to teacher effects models substantially increase the importance of these effects, compared to student background effects? If so, this finding would suggest the importance of including classroom practice variables in future large-scale studies of reading. As will be seen, difficulties encountered in doing this exploratory study indicate that large-scale studies of classroom practices, while vital, raise many methodological hurdles that subsequent research will need to overcome.

To answer these questions, it was necessary to obtain large-scale data that were representative of a large population and included measures of student reading comprehension, teacher background, classroom practices and student and school background. Fortunately, a recent administration of the National Assessment of Educational Progress met these criteria. NAEP was administered to a nationally representative sample of 7,194 fourth graders in 2000 to assess their reading comprehension skills. In addition to the assessment, questionnaires were administered to students and their reading teachers, generating information on their backgrounds and classroom practices. (For overview of the NAEP 2000 Reading Assessment, see National Center for Education Statistics, 2001).

The use of the NAEP, however, introduces certain methodological hurdles that the current study needed to overcome. First, the study needed to appropriately handle variability in the reading comprehension measure. To limit the amount of time students were assessed, each student answered only a limited number of test items; consequently, it is not possible to generate a single student score. Instead NAEP provides five scores based upon the items the student answered and student and school background information. The recommended procedure for conducting secondary analyses using these five scores, known as plausible values, is to estimate a separate model for each and then pool them. The unstandardized and standardized coefficients are pooled by calculating their means and variances through the following formula:

$$v = u + (1.2)B,$$

where v is the pooled variance, u is the average sampling variance and B is the variance among the five plausible values.

Second, the study needed to appropriately handle the sample design. Because NAEP is a clustered, stratified sample, student and teacher observations are not independent of one another. If treated as a simple random sample, these observations will underestimate standard errors. Consequently, standard errors need to be adjusted. One acceptable technique for doing so is using a design effect. NAEP provides weights, known as jackknife weights, that can be used to estimate the effect of the sample design on the standard error of each coefficient, known as the design effect. Because of the computationally expensive nature of estimating the effect for every coefficient in a model, it is appropriate to estimate effects for a subset of coefficients and then select one of these for the purpose of inflating the standard errors of all coefficients.

To relate the teacher and student characteristics to student reading comprehension, the statistical technique of structural equation modeling (SEM) was employed. Like regression analysis, SEM makes it possible to relate independent variables to dependent variables, taking into account both the independent variables and statistical controls. It has two advantages over regression. First, it can test the fit of entire path models, meaning that it can estimate the coefficients and overall goodness of fit of models that relate independent variables to one another as well as to the dependent variable. This makes it possible to incorporate intervening variables into the model. Second, it can construct its independent and dependent variables from observed variables through factor models. This makes it possible both to take into account measurement error and to reduce such error through the use of multiple indicators. (Note 1)

The current study estimated two sets of factor and path models: The first set consisted of five versions of a teacher background model, one for each plausible value, and the second set consisted of five versions of a classroom effects model, also one for each plausible value. These models were estimated using AMOS 3.6 (Arbuckle, 1996), an SEM package, and STREAMS 1.8 (Gustafsson & Stahl, 1997), a pre-and post-processor for SEMs. The factor model portion of the teacher background model constructed measures of four teacher background characteristics (major, education level, years of experience, and perceived preparedness to teach), two student background characteristics (socio-economic status (SES) and home reading behavior), one school characteristic (class size) and one student outcome (a plausible value for reading comprehension performance). SES was constructed from five measures, home reading behavior from four, and the rest from single measures. (Factor models for single measures fix factor loadings at 1 and error terms at 0. See Table 1 for full list of the measures employed.)

Table 1. Descriptive Statistics for Teacher, Student and School Background Characteristics

	M	SD	N
Teacher Background			
Preparedness	-.03	.86	7914
Major in English (1=yes)	.19	.42	7914
Major in Reading/Language Arts (1=yes)	.22	.46	7914
Education Level (1=Master's or more)	.38	.44	7914
Years of Experience (From 1= low to 5=high)	3.33	1.22	7914
Student Socio-economic Status			
Family Subscribes to Newspaper (1=yes)	.72	.43	7914
Family Subscribes to Magazines (1=yes)	.72	.41	7914
Family Owns Encyclopedia (1=yes)	.80	.37	7914
Family Owns More Than 25 Books (1=yes)	.93	.25	7914
Mother College Graduate (1=yes)	.71	.36	7914
Student Reading Background			
Uses Libraries (From 1= never or hardly ever to 4=most every day)	2.60	.93	7914

Talks to Peers about Reading (From 1=never or hardly ever 4=almost every day)	2.64	1.10	7914
Reads for fun (From 1=a poor reader to 4= almost every day)	3.03	1.04	7914
Self-assessment of Reading Skills (From 1=a poor reader to 4=a very good reader)	3.20	.83	7914
School Background			
Class Size (From 1=36 or more students to 5=1-20 students)	4.00	.86	7914

The path portion of the model then related the student outcome to the teacher background measures, taking into account the student background and school measures. The factor portion of the classroom practices model constructed measures of eight classroom practices (teaching metacognitive skills, integrating reading and writing, use of reading materials such as trade books and basal readers, time spent reading in class, working in groups, parental involvement, authentic assessment and traditional assessment), two students background characteristics (SES and home reading behavior), one school characteristic (class size) and one student outcome (a plausible value). Metacognitive skills was constructed from five measures, integrating reading and writing from four measures, reading materials from five measures, time spent reading from two measures, group work from two measures, authentic assessment from four measures and traditional assessment from three measures. (See Table 2 for full list).

Table 2. Descriptive Statistics for Classroom Practices

	M	SD
Metacognitive Skills		
Describe Author's Method (From 1=never or hardly ever to 4=almost every day)	2.61	.80
Explain Reading (From 1=never or hardly ever to 4=almost every day)	3.39	.61
Make Generalizations (From 1=never or hardly ever to 4=almost every day)	3.44	.59
Predict Outcomes (From 1=never or hardly ever to 4=almost every day)	3.48	.59
Learning New Words in Context (From 1=never or hardly ever to 4=almost every day)	3.78	.39
Writing in Service of Reading		
Writing about Literature (From 1=never or hardly ever to 4=almost every day)	.43	.46
Reading and Writing (From 1=never or hardly ever to 4=almost every day)	.70	.43
Writing about Reading (From 1=never or hardly ever to 4=almost every day)	3.18	.69
Answers Questions in Writing (From 1=never or hardly ever to 4=almost every day)	3.37	.61
Reading Materials		
Trade Books (From 1=never or hardly ever to 4=almost every day)	.20	.37
Basal Readers (From 1=never or hardly ever to 4=almost every day)	.17	.34
Reading Kits (From 1=never or hardly ever to 4=almost every day)	1.88	1.01
Children's Newspapers (From 1=never or hardly ever to 4=almost every day)	2.08	.79
Worksheets (From 1=never or hardly ever to 4=almost every day)	2.97	.82

Time Reading		
Reading Aloud (From 1=never or hardly ever to 4=almost every day)	3.59	.57
Reading Silently (From 1=never or hardly ever to 4=almost every day)	3.78	.45
Group Work		
Work in Small Groups (From 1=never or hardly ever to 4=almost every day)	.69	.42
Engage in Group Activities (From 1=never or hardly ever to 4=almost every day)	2.34	.73
Parental Involvement		
Parents Check Homework (From 1=never or hardly ever to 4=almost every day)	2.33	1.01
Parents Sign Off (From 1=never or hardly ever to 4=almost every day)	2.96	.92
Authentic Assessment		
Portfolios (From 1=never or hardly ever to 4=once or twice a week)	1.99	.97
Paragraphs (From 1=never or hardly ever to 4=once or twice a week)	3.19	.80
Projects (From 1=never or hardly ever to 4=once or twice a week)	2.65	.67
Oral (From 1=never or hardly ever to 4=once or twice a week)	3.00	.88
Traditional Assessment		
Multiple-Choice Test (From 1=never or hardly ever to 4=once or twice a week)	2.87	.88
Short-Answer Test (From 1=never or hardly ever to 4=once or twice a week)	3.17	.78
Tests (From 1=never or hardly ever to 4=once or twice a week)	2.49	.67

For the final version of the models, some of the multiple indicator constructs were turned into single indicator constructs to identify which indicator was responsible for the classroom effect of a construct. Thus, integrating reading and writing and reading materials were divided into their constituent indicators. Student background, school characteristics, and student outcomes were measured as per the teacher background models. The path portion of the classroom effects model related the classroom practice constructs to the student outcome, taking into account student background and school characteristics, as well as relating the student background and school characteristics to each of the classroom practices, thus making it possible to gauge the extent to which classroom practices acted as intervening variables between student background, school characteristics and student outcomes. (Note 2)

Results

The factor models and goodness-of-fit statistics reveal that the models fit the data well. For all factor models, the constructs loaded substantially and on all of the corresponding indicators, and all loadings were statistically significant at the .05 level. (Factor models are not presented here, but are available upon request.). All ten of the factor and path models also had adequate goodness-of-fit statistics. For the teacher background models, the RMSEAs were at the .03 level, with normed goodness-of-fit indices of .92 and comparative goodness-of-fit indices at .92 and .93, depending upon the plausible value. For the classroom practice models, the RMSEAs were at the .05 level, with both normed and comparative goodness-of-fit indices at .98. These results suggest that the hypothesized models were confirmed by the observed data.

The path models for teacher background reveal only a modest effect of teaching on student reading comprehension (Table 3). The strongest effects come from students, with SES having the largest effect in the model (b=.37) followed by reading background (b =.14). The school control, class size, also had an effect, albeit a modest one (b =.03). Among the five teacher background variables, only one, years of experience, proved statistically significant, with an unstandardized coefficient of .05. This findings differs somewhat from the literature, in which teacher major tend tends to have an effect and teacher experience tends not to have one. This divergence may be attributable to the fact that this study is of fourth graders and their elementary school teachers, whereas most of the studies of teacher major are at the high school level. (Note 3)

Table 3. Structural Equation Model of Teacher Background Effects

	Reading Achievement
Student Socio-economic Status	140.04** .37
Student Reading Background	16.99** .14
Class Size	1.50** .03
Teacher Preparedness	.72 .02
Teacher Major	1.96 .02
Teacher Education Level	-2.57* -.03
Teacher Experience	1.48** .05
Error	.90

*p <.10
**p<.05

The classroom practice path models reveal much more substantial teacher effects (Table 4). As with the teacher background model, the strongest single effect is of SES (b=.43). This is followed, however, by two teacher effects, the positive effect of metacognitive skill instruction (b= .31) and the negative effect of time spent reading in class (b=.30). Student reading background is next in importance, with students with stronger backgrounds scoring higher on the reading comprehension assessment (b=.13). Teachers' having students write about literature they are reading and using trade books as their primary reading materials had modest positive effects (b=.04 for each). Class size, as in the teacher background model, also had a statistically significant effect of that size.

Table 4. Structural Equation Models of Classroom Effects

	Reading Achievement
Student Socio-economic Status	138.66** .43
Student Reading Background	15.96** .13
Class Size	1.64** .04
Writing	.95** .04
Basal Reading	.19 .01
Trade Books	1.43** .04
Metacognition	1.67* .31
Time Reading	1.38* .30

*p <.10
**p< .05

In addition, the classroom practice path models indicate that students are exposed to very different practices depending upon their background characteristics and those of their schools. Affluent students are more likely to be exposed to metacognitive instruction (b=.04), writing about literature (b=.07) and reading trade books (b=.07) than their less affluent peers. There is, however, no difference in time spent reading in class or the use of basal readers between the two groups. Schools with smaller classes also differ from those with larger classes, with small class students more likely to be exposed to metacognitive instruction (b =.03) and writing about literature (b =.06) as well as to spend more time reading in class (b=.05). It thus appears that effective classroom practices act as intervening variables between student SES and reading comprehension performance, with higher SES students more likely to be exposed to those practices that are themselves associated with higher NAEP scores. With class size, the pattern is less clear, as the practices associated with smaller class sizes may or may not have a positive relationship to NAEP scores.

These findings answer the first research question in the negative and the second in the affirmative. The large-scale data do seem to confirm some of the findings from small-scale research but not others. Some practices, namely metacognition, using trade books and a measure of integrating reading and writing, did prove positively related to reading comprehension. Other practices, however, such as having students work in groups, increasing parental involvement, and the use of authentic assessment, did not. And time spent reading in class actually had a negative relationship to student performance. The addition of classroom practices to large-scale models seems to make the overall impact of teachers comparable to that of student background. As with typical production functions, the teacher background model revealed only a single modest teacher effect. The classroom practice model, however, revealed multiple teacher effects, some of them quite strong. The total standardized effect for the four teacher variables (.70) is actually somewhat larger than the total standard effect of the two student background measures (.56).

Conclusions

These findings have significant methodological implications for research on teacher effects on reading comprehension. The finding that some of the classroom practices proved effective while others did not suggest the need for synergy between small-scale and large-scale research. The findings of small-scale, highly internally valid, studies should serve as the basis for large-scale, highly externally valid, studies. Only in this way can it be known if small-scale findings are applicable to large populations. (This does not rule out the possibility that small-scale research can, by itself, provide information about small populations.) The finding that the introduction of classroom practices leads to substantial teacher effects suggests the need for large-scale research to embrace such variables. Clearly, the failure of previous large-scale research to uncover substantial teacher effects is in large part due to its not including such variables. In addition, the other methodological advances of the current study over traditional production functions proved useful. The use of multiple indicators improved the quality of the measures employed, and the use of path models led to the finding that classroom practices act as intervening variables between student background and reading comprehension performance.

Yet while the current exploratory study does take some steps to improve the large-scale methodology for the study of teacher effects, much remains to be done. One shortcoming of the current study is the ad hoc manner in which it addressed problems with teacher self-reports. Because it relied on pre-existing data, the study made use of interaction effects to increase the likelihood that teachers reporting the use of certain practices were actually using them. Doing so, however, truncated the sample, and is based on the assumption that the more experienced, better prepared teachers are more likely to accurately assess and report what their practices are. This assumption may or may not hold true for a given teacher. A more effective technique for reducing problems with teacher self-reports would be to begin to design questionnaire items that make clearer what the practices are and minimize the social desirability effects. For instance, a questionnaire might include a scenario in a classroom and ask the respondent to describe how he or she would address it. Respondents could also be asked to rank order the effectiveness of the classroom practices of others, or to draw up a time budget for various practices. Such methods, often employed in small-scale research, need to be applied on a larger scale to enhance the reliability and validity of the large-sample teacher reports.

Another shortcoming of the study is its use of cross-sectional data. Because the data are cross-sectional, it is not clear whether particular practices enhance reading comprehension or high performing students are more likely to have teachers engaging in such practices. The study did address this problem in an ad hoc fashion by controlling for measures of student home reading behavior. Indeed, those controls may have resulted in underestimates of teacher effects, in that teachers may positively influence home reading behavior. Whatever the impact of the ad hoc procedure, it is no substitute for longitudinal data that follow student performance over time, and hence it is crucial for subsequent large-scale studies to collect such data. Indeed, the Early Childhood Longitudinal Study (ECLS), which will follow a national sample of students from kindergarten through fifth grade, testing their reading skills and measuring teacher classroom practices, may address this need.

A third shortcoming of this study is its failure to fully take into account the multilevel nature of its data. This study involved multiple levels of analysis in that it related teacher-level inputs to student-level outputs. Yet students are not selected at random, but are clustered within classrooms. The employment of design effects addressed this issue somewhat by increasing standard errors based upon clustering at the level for the school district. But it did not take into account the impact of classroom level non-independence on standard errors. It also did not distinguish between student-level and contextual effects; the influence of student SES, for instance, may be in part due to the average SES of that student's peers. To fully address all of these issues, multilevel techniques, such as Hierarchical Linear Modeling or multilevel versions of SEM (MSEMs) need to be employed.

Finally, rich data is needed on teacher background. This study used the same kinds of summary measures employed by production function studies, such as years of experience and education levels. It may be that teacher background is extremely important, but can only be fully gauged through learning in a more nuanced way about the background. The education level of the teacher may not be important, but the nature and extent of the teacher education curriculum may be. Perhaps certain kinds of induction experiences are more conducive to high student performance. And the nature and extent of professional development experiences may play a role in encouraging particular effective classroom practices. Thus, while the current study suggests that failure to consider classroom practices has led large-scale research to underestimate teacher effects, it may be that the effects of teacher background have also been underestimated. The only way to gauge the full impact of teachers is to collect as much information about them as is collected about their students, and see how the biographies and classroom actions of the two actors unfold together.

In sum, it should be possible to gauge the effectiveness of instructional practices in the area of reading comprehension using large-scale data. The ability to generalize from smaller to larger populations is critical given the existence of new policies that call for "scientifically-based" instruction in reading, that are national in scope. Without knowing the applicability of particular techniques to all populations, policy makers and educators run the risk of imposing that technique on an inappropriate population. Consequently, research should build upon the methods identified in this exploratory study to determine which instructional practices are help for all students, and which are helpful for particular subpopulations of students.

Notes

1. SEM accomplishes this through three steps. First, factor and path models are specified by the researcher. Factor models indicate which observed variables load on which constructs. Path models indicate which constructs are permitted to be related to one another. Second, through an iterative process, the covariance matrix that these specifications imply (S) is matched with a covariance matrix of the observed data (S) to maximize their fit with one another. Finally, the resulting output consists, for each construct, of standardized factor loadings and standard errors for each indicator, standardized and unstandardized path coefficients and their standard errors for each relationships between constructs; and goodness-of-fit statistics including the root mean squared error of approximation (RMSEA) and indices such as the comparative and normed fit indices.
2. Because of issues in using teacher self-reports, the classroom practice measures had to be transformed for the five models. Research has found teacher self-reports of classroom practices to be frequently unreliable. Some teachers may misrepresent the practices in which they engage because they do not fully understand what the named practices are, and some may misrepresent the practices because they perceive the practices as socially desirable. The NAEP data indicated that most teachers claim to engage in most practices, and consequently, giving full weight to the responses of all teachers would make it difficult to distinguish between those that actually do and do not engage in a given practice. Instead, practices were weighted by teachers' years of experience and their perception of their preparedness, giving more weight to the responses of the more prepared, more experienced teachers. This was accomplished through calculating interaction effects between each classroom practice and the two teacher background measures, and substituting these for the classroom practice measures in the models.
3. Teachers' education level had a negative coefficient at the .10 level. This counterintuitive effect requires further exploration.

References

Alfassi, M. (1998). Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial reading classes. *American Educational Research Journal, 35*, 2, 309-332.

Anderson, S. A. (2000). How Parental Involvement Makes a Difference. *Reading Improvement, 37*, 2, 61-86.

Arbuckle, J. L. (1997). *Amos users' guide. Version 3.6*. Chicago: Small Waters Corporation.

Bromley. K. & Modlo, M. (1997). Using cooperative learning to improve reading and writing in language arts. *Reading and Writing Quarterly, 13*, 1, 21-35.

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Cantrell, S. C. (1999). The effects of literacy instruction on primary students' reading and writing achievement. *Reading Research and Instruction, 39*, 1, 3-26.

Cohen, D. K. & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*, 2, 294-343.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York. R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.

Cross, D. R. & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology, 80*, 2, 131-142.

Durkin, D. (1978). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly, 14*, 4, 481-

Durkin, D. (1981) Reading comprehension instruction in five basal reader series. *Reading Research Quarterly, 16*, 4, 515-544.

Epstein, J. L. (2001). *School, family and community partnerships: Preparing educators and improving schools.* Boulder, CO: Westview Press.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal of Legislation, 28*(2), 465-498.

Ferguson, R. F. & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding school accountable: Performance-based reform in education*(pp. 265-298). Washington, D.C.: The Brookings Institution.

Goldhaber, D. D. & Brewer, D. J. (1996) Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources, 32*(3), 505-520.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*, 3, 361-396.

Gustafsson, J. E. & Stahl, P. A. (1997). *STREAMS user's guide (Version 1.7).* Mölndal, Sweden: Multivariate Ware.

Guthrie, J. T. (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. *Reading Research Quarterly, 31*, 3, 306-332.

Guthrie, J. T., Schafer, W. D., Von Secker, C. & Alban, T. (2000). Contributions of instructional practices to reading achievement in a statewide improvement program. *Journal of Educational Research, 93*, 4, 211-225.

Guthrie, J. T., Schafer, W. D. & Huang, C.-W. (2001). Benefits of opportunity to read and balanced instruction on the NAEP. *Journal of Educational Research, 94*, 3, 145-162.

Hansen, J. & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology, 75*, 6, 821-829.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Research, 18*, 4, 45-51.

Hanushek, E. A. (1996a). A more complete picture of school resource policies. *Review of Educational Research, 66*, 3, 397-409.

Hanushek, E. A. (1996b). School resources and student performance. In G. T. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 43-73). Washington, D.C.: The Brookings Institution.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*, 2, 141-164.

Hedges, L. W. & Greenwald, R. (1996). Have times changed? The relation between school resources and student performance. In G. T. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success.* (pp. 74-92). Washington, D.C.: The Brookings Institution.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Research, 23*, 3, 5-14.

Kaniel, S., Licht, P., & Peled, B. (2000). The influence of metacognitive instruction of reading and writing strategies on positive transfer. *Gifted Education International, 15*, 1, 45-63.

Knapp, M. S. (1995).*Teaching for meaning in high-poverty classrooms.* New York, NY: Teachers College Press, Columbia University.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*, 1, 29-45.

Mislevy, R. J. (1993). Should "multiple imputations" be treated as "multiple indicators"? *Psychometrika, 58*, 1, 79-85.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125-145.

Mueller, M. E. (1997). Using metacognitive strategies to facilitate expository text mastery. *Educational Research Quarterly, 20*, 3, 41-65.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22*, 3, 399-420.

National Center for Education Statistics. (1996). *High school seniors' instructional experiences in science and mathematics.* Washington, D.C.: U.S Government Printing Office.

National Center for Education Statistics. (2001). *The nation's report card: Fourth-grade reading highlights.* Washington, DC.: U.S

Government Printing Office.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Bethesda, MD: National Inst. of Child Health and Human Development (NIH),.

O'Reilly, P. E., Zelenak, C. A., Rogers, A. M., & Kline, D. L. (1996). *1994 trial state assessment program in reading secondary-use data files user guide*. Washington, D.C.: U.S. Department of Education.

Popplewell, S. R. & Doty, D. E. (2001). Classroom instruction and reading comprehension: A comparison of one basal reader approach and the four-blocks framework. *Reading Psychology, 22*, 2, 83-94.

Rosenshine, B. & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*, 4, 479-530.

Snow, C. E., Ed.; Burns, M. Susan, Ed.; Griffin, Peg, Ed. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: Committee on the Preservation of Reading Difficulties in Young Children.

Topping, K. J. & Paul, T. D. (1999). Computer-assisted assessment of practice at reading: A large scale survey using accelerated reader data. *Reading and Writing Quarterly, 15*, 3, 213-231.

Wenglinsky, H. (2001). The use of multilevel latent growth modeling to measure school effects. In R. Nata (Ed.). *Progress in Education*, 4 (pp. 167-195). Huntington, NY: Nova Science Publishers.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10*, 12.

Wharton-McDonald, R., Pressley, M., & Hampton, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal, 99*, 2, 101-128.

About the Author

Harold Wenglinsky

Center for Educational Leadership
Baruch College School of Public Affairs
17 Lexington Ave, Box C-305
New York, NY 10010
Email: Harold_Wenglinsky@baruch.cuny.edu

Harold Wenglinsky is Associate Professor in the Baruch School of Public Affairs and Research Director of the Center for Educational Leadership. His research interests include the impact of teachers' instructional practices on student learning and maximizing the effectiveness of educational technology, about which he is currently completing a book.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple University of Wisconsin	David C. Berliner Arizona State University
Greg Camilli Rutgers University	Linda Darling-Hammond Stanford University
Sherman Dorn University of South Florida	Mark E. Fetler California Commission on Teacher Credentialing
Gustavo E. Fischman California State Univeristy–Los Angeles	Richard Garlikov Birmingham, Alabama

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 930 times since July 8, 2003

Volume 11 Number 20 **July 8, 2003** **ISSN 1068-2341**

A Multilevel, Longitudinal Analysis of Middle School Math and Language Achievement

Keith Zvoch
Albuquerque (NM) Public Schools

Joseph J. Stevens
University of New Mexico

Citation: Zvoch, K. and Stevens, J. J. (July 8, 2003). A multilevel, longitudinal analysis of middle school math and language achievement. *Education Policy Analysis Archives, 11*(20). Retrieved [date] from http://epaa.asu.edu/epaa/v11n20/.

Abstract

The performance of schools in a large urban school district was examined using achievement data from a longitudinally matched cohort of middle school students. Schools were evaluated in terms of the mean achievement and mean growth of students in mathematics and language arts. Application of multilevel, longitudinal models to student achievement data revealed that 1) school performance varied across both outcome measures in both subject areas, 2) significant proportions of variation were associated with school-to-school differences in performance, 3) evaluations of school performance differed depending on whether school mean achievement or school mean growth in achievement was examined, and 4) school mean achievement was a weak predictor of school mean growth. These results suggest that assessments of school performance depend on choices of how data are modeled and analyzed. In particular, the present study indicates that schools with low mean scores are not always "poor performing" schools. Use of student growth rates to evaluate school performance enables schools that would otherwise be deemed low performing to demonstrate positive effects on student achievement. Implications for state accountability systems are discussed.

With the enactment of the No Child Left Behind (NCLB; No Child Left Behind Act, 2002) legislation, states are now required to develop content-based standards in mathematics and reading or language arts and have tests that are linked to those standards in grades 3 through 8. The new legislation also requires states to set a proficiency standard for performance on those content-aligned tests. The proficiency standard will enable states to identify "probationary" schools, monitor their performance, and intervene if adequate yearly progress toward the standard does not occur. The increased emphasis that NCLB brings to the assessment of state content standards and the measurement of school effectiveness is intended to ensure that all students have access to an equitable and comprehensive education. However, the assessment of student performance and the measurement of school effectiveness are neither simple nor straightforward (Linn, 2000; Stevens, 2000). There are many complex issues involved in the development and implementation of accountability systems that are not acknowledged or considered in the fervor of political and public dialogue and policy discussion on educational reform. One issue of substantial importance is how the analytic methods used in an accountability system may impact the evaluation of school effectiveness. At its heart, the measurement of student learning and school effectiveness poses some challenges in research design that must be met if the effects of teachers and schools are to be validly estimated.

One of the most difficult challenges in evaluating school performance is to separate the effects of schooling from the intake characteristics of the students who attend the school (Raudenbush & Willms, 1995; Willms, 1992). The evaluative challenge stems from the manner in which students come to attend particular schools. Nonrandom selection processes sort families into neighborhoods and students into schools. The unequal distribution of student characteristics that follows tends to give schools with challenging intakes a competitive disadvantage in most accountability systems. Schools with disadvantaged intakes are at particular risk of unfavorable evaluation if the state accountability system fails to use statistical methods that properly account for student background and the hierarchical nature of school accountability data. State accountability models that use school means or medians as the primary or only indicator of school effectiveness are particularly problematic. Common practice is to aggregate student data to the level of the school in these models. However, if relevant data occur at different levels or for different sampling "units" as in the measurement of both students and schools, then aggregating data to the school level may be inappropriate. This issue is often referred to as a "unit of analysis" problem. In statistical terminology, students are nested within schools and analysis should incorporate the nested structure of the data into the design through the use of multilevel analysis methods (see Aitken & Longford, 1986; Brown, 1994; Burstein, 1980; Cronbach, 1976; Cronbach & Webb, 1975; Goldstein, 1988; Raudenbush & Willms, 1995). Yet, few state systems appear to use multilevel methods. Accountability systems that fail to properly model the nested structure of data tend to confound school intake with school practice and policy and are probably biased in their estimation of school effects.

Another issue of importance in the measurement of school performance is the over-reliance in accountability systems on the comparison of successive cohorts of students as a measure of "change". States that use the successive cohort approach (e.g., the mean performance of 6th graders in 2001 is compared to the mean performance of 6th graders in 2002) to measure and evaluate school performance attempt to mitigate school differences in student intake by focusing on the year-to-year change in student achievement scores. The comparison of successive student cohorts enables states to evaluate schools in terms of proficiency gains instead of absolute performance levels. However, the use of different cohorts of students to measure school progress or school improvement is problematic for evaluative purposes. Recent investigations of the successive cohort approach demonstrate that estimates of year-to-year gains in proficiency are affected in large part by sampling variation, measurement error, and unique, non-persistent factors that are not associated with school size or school practice (Linn & Haug, 2002; Kane & Staiger, 2002). The lack of systematic variation in the successive cohort change score puts states at risk of assessing school performance on the basis of fluctuations in student cohorts or test administration conditions instead of actual changes in student performance (Linn & Haug, 2002).

Evidence that school performance cannot be estimated without bias when student test scores are aggregated at a single point in time or with precision when successive student cohorts are compared has led a number of authors to argue for the use of longitudinal analyses of individual student performance as a more direct and accurate estimate of school effects (Barton & Coley, 1998; Bryk & Raudenbush, 1988; Linn & Haug, 2002). For example, Goldstein (1991, p. 14), describing school effectiveness studies in Britain, stated that "...It is now recognised...that *intake* achievement is the single most important factor affecting subsequent achievement, and that the only fair way to compare schools is on the basis of how much progress pupils make during their time at school." Student progress can be measured by comparing year-to-year differences in individual performance, but the most appropriate methodology for measuring changes in student achievement is through estimation of individual growth trajectories by means of the multilevel model (Bryk & Raudenbush, 1992, 1987; Willett, 1988; Willms, 1992). In this approach, student test scores are linked across time. A regression function is then fit to the outcome data obtained on each student. The resulting growth trajectories index the rate at which students acquire certain academic competencies. A measure of school performance follows from averaging the individual growth trajectories within each school.

Multilevel, longitudinal analyses of individual student performance may allow the conceptualization of the most relevant and direct outcome measures of school effectiveness by facilitating estimation of the added benefit or "value" that students receive by attending a particular school (Boyle & Willms, 2001; Bryk & Raudenbush, 1988; Willms, 1992). The multilevel, longitudinal model facilitates value-added school performance estimates by providing a degree of control over a wealth of confounding factors that otherwise complicate the evaluation of school effectiveness. When longitudinal models are used, each student serves as his/her own control for confounding factors that are stable characteristics of the student over time (Sanders & Horn, 1994; Stevens, 2000). Therefore, the confounding effects of factors like socio-economic status, limited English proficiency, and ethnic and cultural differences may be largely controlled through the application of a matched longitudinal design.

Despite the potential of using multilevel, longitudinal models to measure and evaluate school performance (Boyle & Willms, 2001; Teddlie & Reynolds, 2000), only a few reported studies have applied these models to student achievement data (e.g., Bryk & Raudenbush, 1988; Willms & Jacobsen, 1990). Given the lack of published examples, the purpose of the present study was to provide a demonstration of the use of multilevel, longitudinal models to estimate school effectiveness using a sample of middle school students. We were interested in examining the following research questions: 1) How does student achievement performance vary over a three year period? 2) How much of the variation in performance is associated with individual differences among the students and how much with differences from school to school? and, 3) How does the evaluation of school performance differ based on an examination of school mean achievement vs. an examination of school average rates of growth in achievement?

Method

Participants

Standardized test data from middle school students in a large urban school district located in the southwestern United States were analyzed in the present study. The school district that provided the data has over 100 schools and serves close to 90,000 students annually. The district has a diverse student body. In recent years, the student population has been approximately 46% Hispanic, 44% Anglo, 4% Native American, 3% African American, 2% Asian and 1% other. The district serves many students who are not fully English proficient. On average, twenty percent of the students who attend district schools are classified as Limited English Proficient (LEP). The district is also impacted by widespread poverty. In any given year, approximately 35% of the district's middle school students receive a free or a reduced price lunch.

At the middle school level, the school district has 24 schools that serve over 20,000 students in grades 6 through 8. All sixth, seventh, and eighth grade students are tested annually on a norm-referenced achievement test, the TerraNova/CTBS5 Survey Plus (CTB/McGraw-Hill, 1997). Approximately 6,500 students in each grade take the test each spring. Achievement data from students who were in sixth grade in 1998-99, seventh grade in 1999-00, and eighth grade in 2000-01 were analyzed in the present study. Middle school students were used because they provided the only cohort on which three consecutive years of data were available. Mathematics and Language scores were used to provide a demonstration using core subject areas and those content areas required in the NCLB legislation. All students who completed an examination in all three study years were selected (*N* = 4,918). Since the purpose of the study was to examine school effects, 800 students who did not attend the same middle school in all three years were excluded resulting in a sample of 4,118 students. Any student who did not have a mathematics or language composite score in all three study years was excluded as well as any student who received a modified test administration in any of the three years. This resulted in a final sample of 3,299 students nested within 24 middle schools.

The sample consisted of almost equal numbers of males and females. Fifty-one percent of the sample were female (*N* = 1,698); forty-nine percent were male (*N* = 1,601). Representation of ethnic groups was more variable. Forty-six percent (*N* = 1,524) of the sample were Anglo, 45% (*N* = 1,495) were Hispanic, 3% (*N* = 87) were African American, 3% (*N* = 86) were Native American, and 2% (N = 67) were of Asian descent. The ethnic background of 40 students was not identified. Thirty-five percent (*N* = 1,152) of the sample received a free or a reduced price lunch, 12% (*N* = 390) were classified as LEP, and 3% (*N* = 98) were special education students. In most respects, the backgrounds of students in the analytic sample were representative of the students who attend district middle schools. However, the exclusion of students who did not participate in all three test administrations, students who transferred schools, and students who received at least one modified test administration did lower the percentage of free lunch recipients and the percentage of LEP and special education students below district averages by 1%, 8%, and 18%, respectively. Nonetheless, the disproportionate exclusion of students from special populations did not affect the pattern of school mean achievement. Correlations between school mean achievement in grade 6 for the original and the analytic sample were .98 for mathematics and .99 for language.

Measures

Achievement data used in the study were student scores on the TerraNova/CTBS5 Survey Plus, a standardized, norm referenced achievement test (CTB/McGraw-Hill, 1997). The Survey Plus is a test battery that spans grades 2 through 12. All test items are selected-response. The Survey Plus tests students in Reading, Language Arts, Mathematics, Science, Social Studies, Word Analysis, Vocabulary, Language Mechanics, Spelling, and Mathematics Computation. CTB/McGraw-Hill calculates an IRT derived score for each student in each subject area. CTB/McGraw Hill also provides each student with a weighted composite score in Reading, Mathematics, and Language.

Student scale scores on the Mathematics and Language composites were analyzed in the present study. The mathematics composite score is derived from the 31-item Mathematics and the 20-item Mathematics Computation subtests. According to the publisher, the Mathematics subtest measures a student's ability to apply grade appropriate mathematical concepts and procedures to a range of problem-solving situations. The Mathematics Computation subtest measures a student's ability to perform arithmetic operations on grade appropriate number types. CTB/McGraw-Hill reported a KR-20 reliability estimate of .86 for the Mathematics subtest in the 6th, 7th, and 8th grade norming samples. For Mathematics Computation, KR-20 was .83 in grade 6, .80 in grade 7, and .85 in grade 8. For the Mathematics composite, KR-20 was reported at .91 in grade 6, .90 in grade 7, and .92 in grade 8 (CTB/McGraw-Hill, 1997).

The Language composite was also derived from a weighted combination of two subtests. The 22-item Language subtest is intended to measure a student's ability to understand the structure and usage of words in standard written English. The 20-item Language Mechanics subtest is designed to measure a student's ability to edit and proofread standard written English. CTB/McGraw-Hill reported KR-20 reliability estimates of .86 in grade 6, .84 in grade 7, and .81 in the grade 8 norm groups. For Language Mechanics, KR-20 was reported as .84 in grades 6 and 7 and .85 in grade 8. For the Language composite, KR-20 was .91 in grades 6 and 7 and .90 in grade 8 (CTB/McGraw-Hill, 1997).

Analytic Procedures

Multilevel modeling techniques were used to estimate a mean achievement score and a mean growth trajectory for each school. The Hierarchical Linear Modeling (HLM) program, version 5.04 (Raudenbush, Bryk, Cheong, & Congdon, 2001) was used to estimate three-level longitudinal models. Level-1 was composed of a longitudinal growth model that fitted a linear regression function to each individual student's achievement scores over the three years studied (grades 6, 7, and 8). Equation 1 specifies the level-1 model:

$$(1) \qquad Y_{tij} = \pi_{0ij} + \pi_{1ij}(Year) + e_{tij}$$

As written, y_{tij} is the outcome (i.e., mathematics or language achievement) at time t for student i in school j, π_{0ij} is the initial status of student ij (i.e., 6th grade performance), π_{1ij} is the linear growth rate across grades 6-8 for student ij, and e_{tij} is a residual term representing unexplained variation from the latent growth trajectory. Levels 2 and 3 in the HLM model estimate mean growth trajectories in terms of both initial status and growth rate across students (equations 2a and 2b) and across schools (equations 3a and 3b):

$$(2a) \qquad \pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$(2b) \qquad \pi_{1ij} = \beta_{10j} + r_{1ij}$$

$$(3a) \qquad \beta_{00j} = \gamma_{000} + u_{00j}$$

$$(3b) \qquad \beta_{10j} = \gamma_{100} + u_{10j}$$

The initial status and growth of student achievement in equations 2a and 2b is conceived as a function of the school average achievement or school average slope and residual. Similarly, the initial status and growth by school in equations 3a and 3b is conceived as a function of the grand mean achievement or the grand mean slope and residual. Equations 3a and 3b were used to calculate estimates of school mean achievement and school mean growth reported in the present study.

Results

Model Assumptions

Visual examination of univariate frequency distributions and a check of summary statistics revealed that mathematics and language achievement scores were distributed normally (i.e., skew and kurtosis values < 1) in all three study years. A check of within-subject bivariate plots revealed linear relationships between achievement scores across the three study years in both mathematics and language. After checking model assumptions, three SPSS data files were transferred to the HLM program for analysis. The Level-1 data file contained student and school identifiers, three years of student mathematics and language composite scale scores, and a field for year. This file contained 9,897 records (i.e., three records for each of 3,299 students). The Level-2 data file contained student and school identifiers (N = 3,299). The Level-3 data file contained only a school identifier (N = 24).

Mathematics

Table 1 presents the results of the three-level HLM model for mathematics. In the upper portion of the table, the results of the fixed effects regression model are presented. The first estimate shown, the grand mean (γ_{000}), is the intercept or the average 6th grade mathematics scale score for all students in the sample. The second estimate, the grand slope (γ_{100}), is the average yearly growth rate of the students. Thus, in this sample, the average mathematics score is estimated as 659.43 and on average, student mathematics achievement is expected to increase by 18.40 scale score points per year.

Table 1
Three-Level Unconditional Model for Mathematics Achievement

Fixed Effect	*Coefficient*	*SE*	*t*

School Mean Achievement, γ_{000}	659.43	2.97	222.20*
School Mean Growth, γ_{100}	18.40	0.93	19.86*
Random Effect	*Variance Component*	*df*	*Chi-square*
Individual Achievement, r_{0ij}	766.00	3275	8828.40*
Individual Growth, r_{1ij}	26.66	3275	3791.30*
Level-1 Error, e_{tij}	310.87		
School Mean Achievement, u_{00j}	203.03	23	704.70*
School Mean Growth, u_{10j}	19.11	23	359.31*
Level-1 Coefficient		*Percentage of Variation Between Schools*	
Individual Achievement, π_{0ij}		21.0	
Individual Growth, π_{1ij}		41.8	

Note. Results based on data from 3,299 students distributed across 24 middle schools.

* $p < .001$

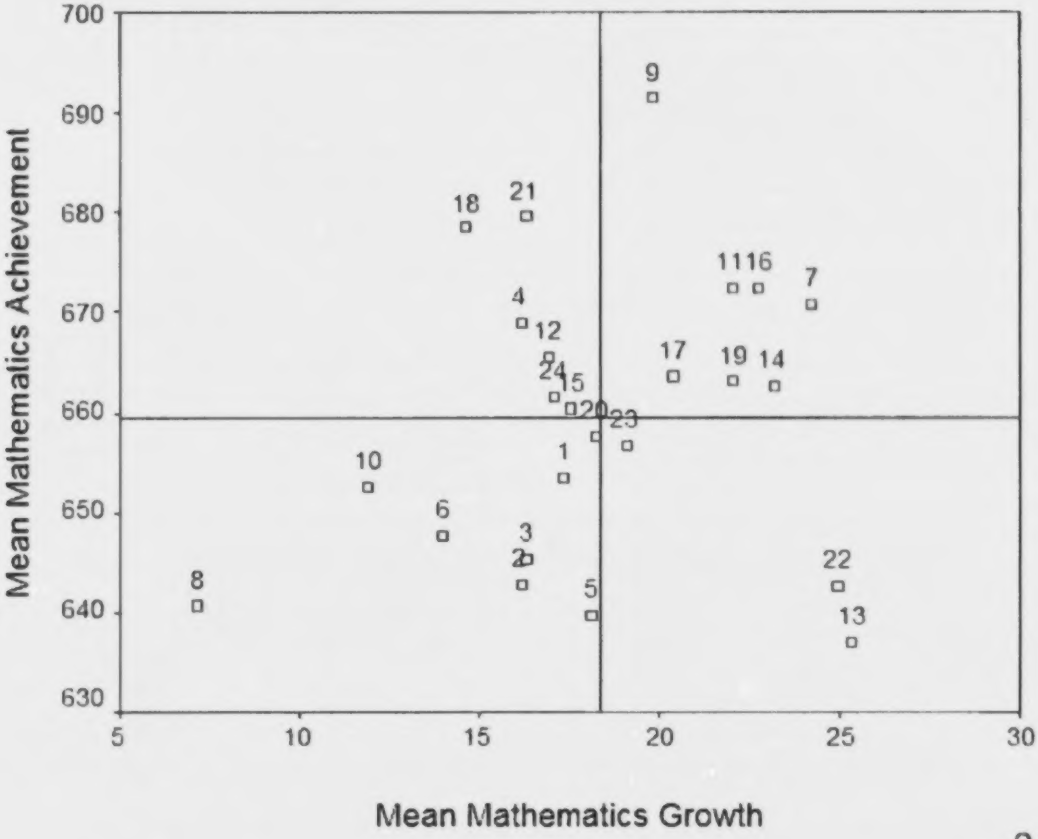


Figure 1. Relationship between mathematics achievement and mathematics growth by school.

Estimates of student and school-level parameter variance are presented next in Table 1. Chi-square tests of the hypotheses that students and schools differ in level and growth of mathematics achievement indicate that there was statistically significant variation across all parameters. Both students and schools differ significantly in initial achievement levels and the rate of achievement growth. This indicates that there are individual differences from one student to another in mathematics achievement initially in grade 6 as well as in the rate of growth in mathematics achievement throughout middle school. In addition, inspection of the variance components presented at the bottom of Table 1 show that the amount of between school variance in mathematics mean achievement (21.0%) and mean achievement growth (41.8%) is also relatively large and statistically significant. Thus, over and above individual differences, there are systematic differences from one school to another in mean mathematics achievement initially in grade 6 and in the school average rate of growth in mathematics achievement from the 6th to the 8th grade.

To further illustrate these observed school level differences in mathematics achievement, Empirical Bayes (EB) estimates of the 24 middle-school mathematics mean achievement and mean growth rates are presented in the scatterplot in Figure 1 on the vertical and horizontal axes, respectively. The horizontal line in the interior of the figure represents the grand mean achievement in mathematics. The vertical line in the interior of the figure represents the grand mean growth in mathematics. The two grand mean reference lines are used to classify schools into four quadrants of school performance. The upper right quadrant contains schools with above average mean achievement in grade 6 and above average growth from grades 6 to 8. The lower right quadrant contains schools with below average mean scores, but above average growth. The two quadrants on the left side of the figure contain schools with below average growth and either high or low mean achievement. A number of interesting results can be seen in Figure 1. First, two schools (22 and 13) with low mean scores record the highest growth in the district. Strong growth is also evident in high scoring school 7. Also evident in Figure 1 is a school (8) with a low mean score and very poor mathematics growth. Relatively poor mathematics growth also occurs in two schools with high 6th grade mathematics achievement. Schools 21 and 18, second and third in 6th grade mean achievement, are noticeably below the district average in mathematics growth. Overall, a slight positive relationship exists between mathematics mean achievement and mathematics mean growth ($t_b = .14$). On average, schools with low mean scores record less growth than schools with high mean scores. Appendix A presents the individual school mean and school growth estimates on which Figure 1 is based.

Figure 2 displays the school estimates in growth trajectory form. Each line in Figure 2 shows the average mathematics achievement at one of the 24 middle schools. As can be seen, there is a good deal of variation from school-to-school both in initial status (i.e., grade 6 mean achievement) and in the average rates of growth over time. The variation in average mathematics growth rates is indicated in part by the number of crossing lines in the figure. Alternative line styles are used to highlight schools with exceptionally high or low growth rates. Schools with a high growth rate are represented by the broken dot pattern. Schools with a low growth rate are represented by the broken line pattern.

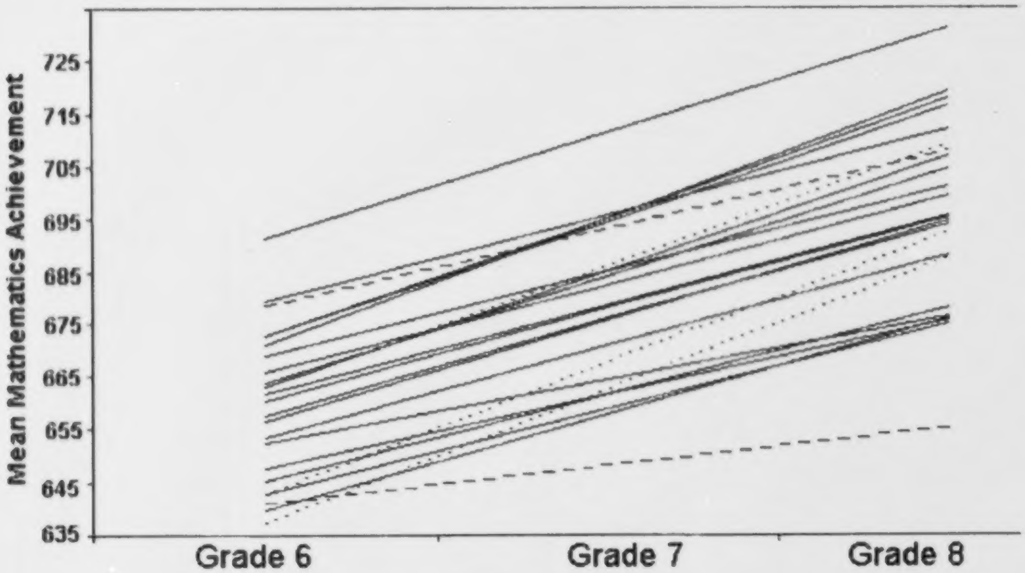


Figure 2. Mean mathematics achievement as a function of grade level and school location.

In Figure 2, the strong growth of two of the schools with low 6th grade achievement levels can be clearly seen. The school with the lowest 6th grade mean score (24th in rank) shows average achievement growth 1.4 times the overall average for mathematics growth. By 8th grade, the rank of the school has changed from 24th to 18th in average mathematics achievement. Similarly, the school that ranks 21st in mean performance in 6th grade, also shows average achievement growth 1.4 times the average and moves to a rank of 16th by the end of 8th grade. Strong growth is also apparent in some of the schools with high 6th grade mean scores. For example, the 10th ranked school in 6th grade mathematics achievement becomes the 6th ranked school in 8th grade mathematics achievement. Schools with lower than average mathematics growth are also readily apparent in Figure 2. The third ranked school in 6th grade mathematics performance falls to seventh ranked in 8th grade performance. In addition, the 22nd ranked school in 6th grade achievement not only becomes the lowest ranked school by the end of 8th grade, but by achieving at only 39% of the overall average for mathematics growth, also falls far behind the achievement level of all other middle schools in the district.

Language

The same three-level, longitudinal HLM model was applied to the language achievement scores of the same sample of students. Table 2 presents these results. As can be seen in Table 2, model results for language achievement were similar to those for mathematics achievement. Except for variation in student growth rates, all parameters of the three-level HLM model were statistically significant. The average language achievement for all students across the 24 middle schools was 661.43 in grade 6 and the average yearly growth in language achievement was 12.30 score points. Inspection of the variance components from the language model shows that while there is statistically significant individual variation in students' initial language achievement in grade 6, individual language growth rates do not differ statistically. Table 2 also shows that school growth rates in language are less variable than school growth rates in mathematics. Of the variation that does exist in language growth rates, 84% was between school variance. Thus, in the case of language achievement, students differed in their average language achievement at grade 6 but showed rates of growth in language achievement that did not differ significantly. At the school level, there were statistically significant differences in average achievement at grade 6 and in average rates of growth in language achievement through grade 8.

Table 2
Three-Level Unconditional Model for Language Achievement

Fixed Effect	*Coefficient*	*SE*	*t*
School Mean Achievement, γ_{000}	661.43	2.58	256.55*
School Mean Growth, γ_{100}	12.30	0.45	27.44*
Random Effect	*Variance Component*	*df*	*Chi-square*
Individual Achievement, r_{0ij}	699.35	3275	8836.36*
Individual Growth, r_{1ij}	0.68	3275	3226.22+
Level-1 Error, e_{tij}	332.11		
School Mean Achievement, u_{00j}	151.66	23	588.34*
School Mean Growth, u_{10j}	3.51	23	91.84*
Level-1 Coefficient		*Percentage of Variation Between Schools*	
Individual Achievement, π_{0ij}		17.8	
Individual Growth, π_{1ij}		83.8	

Note. Results based on data from 3,299 students distributed across 24 middle schools.

*$p < .001$

Empirical Bayes estimates of the 24 middle-school language means and growth rates are displayed in Figure 3. Instances of all four patterns of achievement described for the mathematics achievement results are also present in Figure 3. School 22 demonstrates high growth relative to its 6th grade mean language achievement while growth is low for school 8 in language as it was in mathematics. As with the mathematics results, school 18 again demonstrates low growth relative to 6th grade mean achievement while school 7 shows a high growth rate in language achievement. Overall, the relationship between language mean achievement and language mean growth is positive (t_b = .41). On average, schools with low language mean scores showed less growth than schools with high language mean scores. School language achievement means and growth rate estimates are presented in Appendix A.

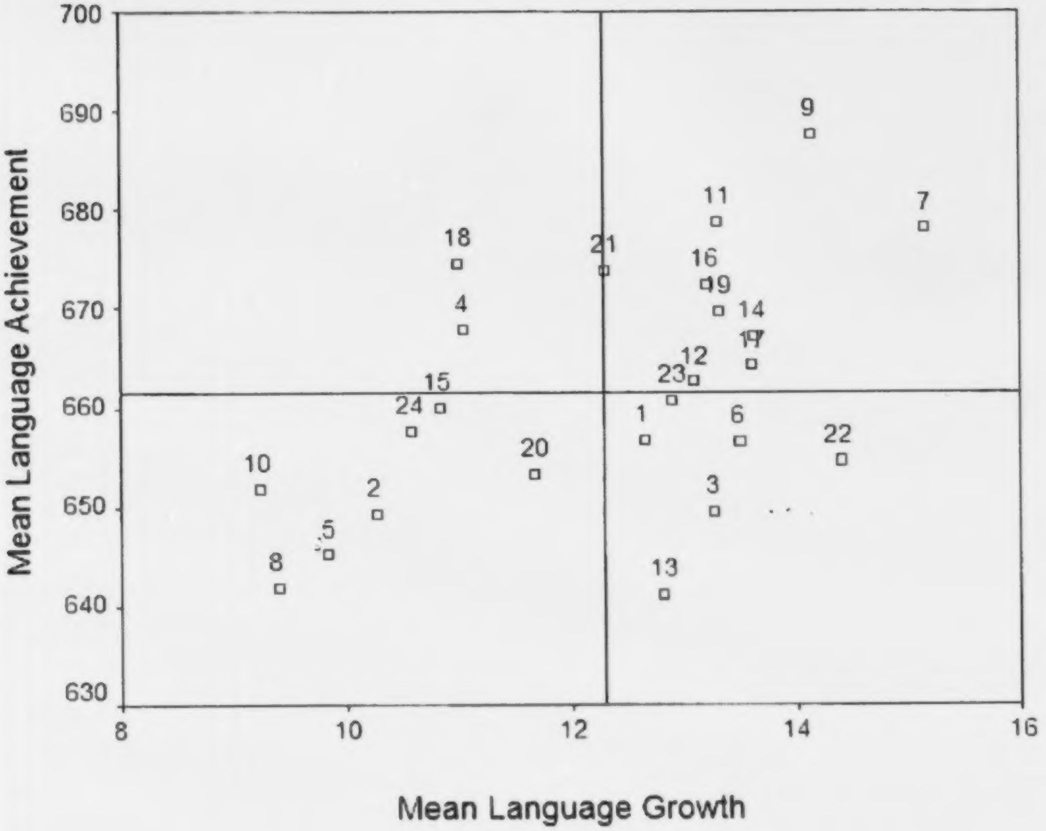


Figure 3. Relationship between language achievement and language growth by school.

Figure 4 displays these results in growth trajectory form. Alternative line styles are again used to represent schools with exceptionally high or low growth rates. The figure shows that, while middle schools differ substantially in mean language achievement at grade 6, the rate of language achievement growth is more similar across schools as evidenced by the parallel pattern of the growth trajectories. Relative to mathematics, fewer schools change rank position. Some possibly important differences in growth rate still exist however. The third ranked school in 6th grade language performance increased its relative standing over other schools in the district. Conversely, the 23rd ranked school in 6th grade language performance becomes the lowest ranked school by the end of 8th grade.

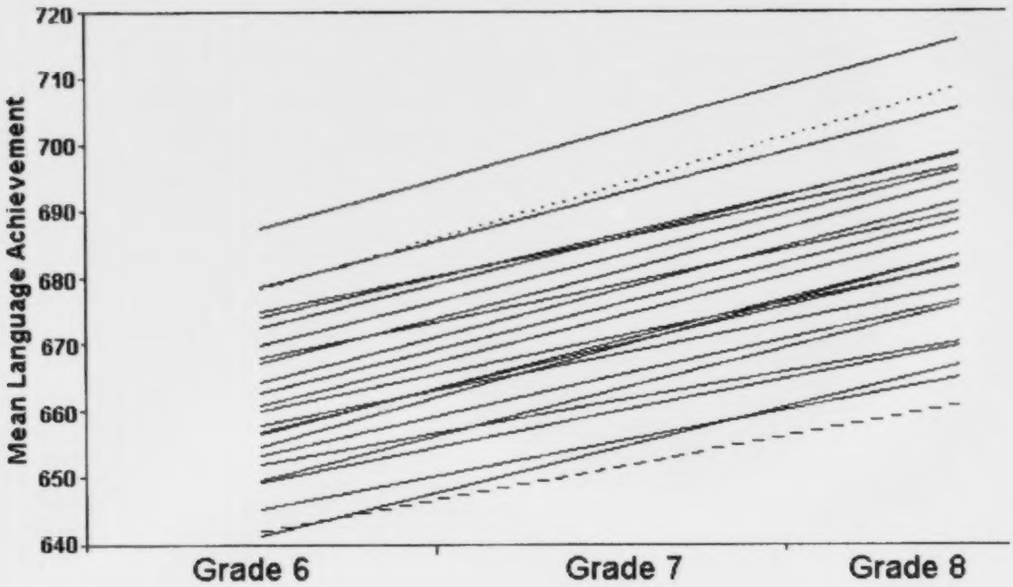


Figure 4. Mean language achievement as a function of grade level and school location.

Discussion

The purpose of the present study was to apply multilevel, longitudinal models in an analysis of school effectiveness in mathematics and language achievement and to demonstrate how assessments of school performance can differ based on how data are modeled and analyzed. The present study demonstrated that assessments of school performance varied across mathematics and language achievement measures. Estimates of the proportions of variance in achievement associated with individual students and with middle schools showed that significant proportions of variation were associated with school-to-school differences in performance. In the current sample, 21 percent of the unadjusted variation in mathematics achievement and 42 percent of the unadjusted variation in mathematics growth were attributable to between-school differences. For language, 18 percent of the unadjusted variation in school achievement means and 84 percent of the unadjusted variation in school growth trajectories were associated with school-to-school differences.

The present study also showed that evaluations of school performance differ depending on whether school mean achievement or school mean growth in achievement are examined. There was significant variation in the mean achievement of students in mathematics and language both from student-to-student and from school-to-school. The analyses also showed that there was significant variation in the rate of achievement growth from student-to-student and from school-to-school for mathematics and from school-to-school for language. Using results from the multilevel, longitudinal models, mean achievement and mean growth were estimated for each middle school in mathematics and language. Evaluation of these estimates showed that the school mean level of performance was not strongly predictive of the school mean rate of growth. Correlations of school mean and school growth estimates were only .14 for mathematics and .41 for language. Inspection of Figures 1 and 3 showed that characterization of school performance is substantially different depending on whether mean achievement or mean growth is examined. In several cases, schools with low mean scores were not always "poor performing" schools. In fact, schools with low mean scores were in many cases the schools with the largest growth rates. Conversely, a high mean achievement score was not always a clear indicator of "good performance". In several instances, schools with high mean scores had growth rates below the district average.

The demonstration that school performance can vary on the basis of the analytic model applied suggests that it is essential to use evaluative models that do not unfairly reward or penalize schools for factors that are beyond the control of school personnel (Hanushek & Raymond, 2001; Ladd, 2001). Current evaluative practice often falls short of this goal. The accountability systems now in use in many states apply evaluative methods that cannot separate school level variation from student level variation or validly disentangle school effects from factors that are outside the control of educational policy and practice at the school (Stevens, Estrada & Parkes, 2000). One of the most common approaches in state accountability systems is the use of the school mean as the only or the key component in the evaluation of school effectiveness. As an evaluative measure, the school mean has widespread appeal. School means are easily calculated and are readily understood. However, the school mean is also a biased indicator of school performance (Heck, 2000). School means reflect all influences on student performance, including those exogenous to the school (e.g., family background, prior achievement, community context). As a result, the school mean often provides a misleading picture of school performance. Schools with advantaged intakes tend to be evaluated more favorably than schools with disadvantaged intakes, regardless of the impact the school has on students over time (Stevens et al., 2000).

Another option for assessing school performance is available to those states or districts that collect comprehensive data on student background. School means can be adjusted on the basis on student characteristics, prior achievement levels, and community characteristics in an attempt to arrive at a mean value that isolates the contribution of school practice and policy (Clotfelter & Ladd, 1996; Raudenbush & Willms, 1995; Willms, 1992). However, these data are often difficult for states and districts to adequately and accurately collect and analyze. Adjusted school means also present states and districts with two unwanted concerns. One concern has to do with public response to having a lower standard of performance for certain special student populations. The second stems from the difficulty of having to convey the meaning of complex statistical adjustments to parents, teachers, and school administrators (Clotfelter & Ladd, 1996; Elmore et al., 1996).

An alternative to the adjusted school mean is a measure based on changes in students' academic achievement over time. As a measure of school performance, school mean growth may offer a more tractable and accurate method of adjusting for socio-demographic characteristics by providing control over confounding influences associated with the stable characteristics of students (Haertel, 1999; Lane & Stone, 2002; Stevens et al., 2000). Repeated measurement of individual students provides control over the background and intake characteristics that strongly impact the *level* at which a student performs by shifting the measurement process from indexing student performance at a single point in time to tracking the rate of pupil progress over time (Sanders & Horn, 1994). The calculation of student growth rates thus enable schools to be evaluated in terms of the gains students make instead of the level at which students start, thereby enabling a more valid comparison of schools that differ in the intake characteristics of their student bodies.

Despite the promise of using multilevel, longitudinal models to measure school performance, very few states have accountability systems that track individual students over time or use analytic methods that account for the hierarchical structure of accountability data (Council of Chief State School Officers, 2001; Education Week, 2001, 2002). However, the importance of basing an accountability system on an outcome measure that can be impacted by school practice and policy cannot be overstated. If school effectiveness is not evaluated in a way that actually reflects school practices and policies but instead reflects student intake characteristics, the state system for evaluating school performance can become a source for flawed decision-making and a target of criticism and possible litigation by disgruntled stakeholders (Parkes & Stevens, in press). Misguided assessment policy can thus stall attempts at constructive school-based change and effectively undermine the intent of the accountability system.

The present study demonstrated that assessments of school performance vary on the basis of the analytic methods applied to the data. Depending on whether schools were evaluated in terms of mean achievement or mean growth, assessments of school performance were shown to differ dramatically. In some cases, schools with low mean scores had high growth rates and schools with high mean scores had low growth rates. These results suggest that states should not rely on the school mean as the sole indicator of school effectiveness. Instead, consideration should be given to incorporating into school accountability systems measures that track student learning or growth over time. The importance of assessing student growth is further underscored by the amount of variation in growth rates that can be attributed to school-to-school differences. In the present study, school differences in mean growth were two times greater than school differences in mean achievement in mathematics and over four times greater than school differences in mean achievement in language. Identification of large amounts of school level variation in growth rates suggests that schools can have substantial influence on student achievement. Future research that examines the influence of school demographic factors, school climate, and school practice and policy on school growth trajectories will begin to facilitate our understanding of why some schools are more effective at promoting student growth in achievement than others.

References

Aitkin, M. & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A, 149*, 1-43.

Barton, P. & Coley, R. (1998). *Growth in school: Achievement gains from the fourth to the eighthgrade.* Princeton, NJ: Educational Testing Service.

Boyle, M.H., & Willms, J.D. (2001). Multilevel modeling of hierarchical data in developmental studies. *Journal of Child Psychology and Psychiatry, 42*, 141-162.

Brown, S. (1994). School effectiveness research and the evaluation of schools. *Evaluation and Research in Education, 8*, 55-68.

Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147-158.

Bryk, A.S., & Raudenbush, S.W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In R.D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 159-204). San Diego: CA: Academic Press.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: SAGE.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D.C. Berliner (Ed.), *Review of research in education* (vol.8). Washington, DC: American Educational Research Association

Clotfelter, C.T., & Ladd, H.F. (1996). Recognizing and rewarding success in public schools. In H.F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 23-63). Washington, DC: The Brookings Institution.

Council of Chief State School Officers (2001).*Annual Survey: State Student Assessmenti Programs, Vols. 1 and 2 (1999-2000 data)*.Washington, D. C.: Council of Chief State School Officers.

Cronbach, L.J. (1976). *Research on classrooms and schools: Formulation of questions, design, andanalysis*. An Occasional Paper, Stanford, CA: Stanford Evaluation Consortium.

Cronbach, L. & Webb, (1975). Between and within class effects in a reported aptitude by treatment interaction: Reanalysis of a study by G.L. Anderson. *Journal of Educational Psychology, 6*, 717-724.

CTB/McGraw-Hill (1997). *TerraNovaTechnical Bulletin 1*. Monterey, CA: Author.

Education Week. (2001). *Quality counts 2001: A better balance*. (Volume 20, Number 17, January 11, 2001). Bethesda, MD: Author.

Education Week. (2002). *Quality counts 2002: Building blocks for success*. (Volume 21, Number 17, January 10, 2002). Bethesda, MD: Author.

Elmore, R. F., Abelmann, C.H., & Fuhrman, S.H. (1996). The new accountability in state education reform: From process to performance. In H.F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 65-98). Washington, DC: The Brookings Institution.

Goldstein, H. (1991). Better ways to compare schools? *Journal of Educational Statistics, 16*(2), 89-92.

Goldstein, H.I. (1988). Comparing schools. In H. Torrance (Ed.). *National assessment and testing: A research response*. London: BERA.

Haertel, E. H. (1999). Validity arguments for high-stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice, 18* (4), 5-9.

Hanushek, E.A., & Raymond, M.E. (2001). The confusing world of educational accountability. *National Tax Journal, 54*, 365-384.

Heck, R.H. (2000). Examining the impact of school quality on school outcomes and improvement: A value-added approach. *Educational Administration Quarterly, 36*, 513-552.

Kane, T.J., & Staiger, D.O. (2002). Volatility in school test scores: Implications for test-based accountability systems. *Brookings Papers on Educational Policy, 1*, 235-283.

Ladd, H.F. (2001). School-based educational accountability systems: The promise and the pitfalls. *National Tax Journal, 54*, 385-400.

Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice, 21*(1), 23-30.

Linn, R.L. (2000). Assessments and accountability. *Educational Researcher, 29*, 4-16.

Linn, R.L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24 (1), 29-36.*

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).

Parkes, J. & Stevens, J. J. (in press). Legal issues in school accountability systems. *Applied Measurement in Education*.

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., & Congdon, R.T. (2001). *HLM 5: Hierarchical linearand nonlinear modeling*. Chicago: Scientific Software International.

Raudenbush, S.W., & Willms, J.D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20*, 307-335.

Sanders, W.L., & Horn, S.P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299-311.

Stevens, J. J. (2000). *Educational accountability systems: Issues and recommendations for New Mexico* (Technical Report). New Mexico State Department of Education.

Stevens, J. J., Estrada, S., & Parkes, J. (2000). *Measurementissues in the design of state accountabilitysystems*. Paper presented at the annual meeting of the American Educational Research Association,

Willett, J.B. (1988). Questions and answers in the measurement of change. In E. Rothkopf (Ed.), *Review of research in education 1988-89* (pp. 345-422). Washington: American Educational Research Association.

Willms, J.D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: FalmerPress.

Willms, J.D. & Jacobsen, S. (1990). Growth in mathematics skills during the intermediate years: Sex differences and school effects.*International Journal of Educational Research, 14*, 157-174.

About the Authors

Keith Zvoch
Albuquerque Public Schools
Albuquerque, NM

Email: zvoch@aps.edu

Keith Zvoch earned a doctorate in Quantitative Methods from the Educational Psychology Program at the University of New Mexico in 2001. He is currently a research scientist for the Albuquerque Public School District. Dr. Zvoch also teaches research methods and statistics at the University of New Mexico on an adjunct basis. His current research interest is the measurement and assessment of school effects.

Joseph J. Stevens
University of New Mexico
College of Education
Educational Psychology Program
Simpson Hall
Albuquerque, NM 87131

Email: jstevens@unm.edu

Joseph J. Stevens is a Professor of Educational Psychology at the University of New Mexico. Dr. Stevens' research concerns applications and validity of large-scale assessment systems, the evaluation of accountability systems and school effectiveness, and the assessment of cognitive diversity.

Appendix A

Sample Sizes and Empirical Bayes Achievement Mean and Slope Estimates by School and Content Area

	Mathematics		Language		
School	Mean	Slope	Mean	Slope	N
1	653.40	17.34	656.67	12.65	71
2	642.78	16.15	649.29	10.26	104

3	645.24	16.32	649.52	13.27	115
4	668.92	16.16	667.71	11.04	165
5	639.67	18.12	645.18	9.83	168
6	647.62	13.98	656.45	13.50	68
7	670.82	24.22	678.12	15.16	178
8	640.74	7.15	641.76	9.39	129
9	691.48	19.83	687.51	14.15	168
10	652.49	11.90	651.85	9.22	88
11	672.49	21.99	678.75	13.31	232
12	665.55	16.95	662.57	13.09	131
13	636.90	25.33	641.20	12.83	103
14	662.64	23.15	667.02	13.62	216
15	660.36	17.55	659.82	10.82	118
16	672.47	22.75	672.41	13.21	242
17	663.63	20.41	664.22	13.61	109
18	678.35	14.64	674.60	11.00	125
19	663.06	22.02	669.66	13.31	172
20	657.59	18.23	653.24	11.69	136
21	679.50	16.27	673.83	12.30	164
22	642.46	24.94	654.66	14.40	97
23	656.64	19.08	660.70	12.91	123
24	661.63	17.07	657.63	10.57	77

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

David C. Berliner
Arizona State University

Greg Camilli
Rutgers University

Linda Darling-Hammond
Stanford University

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Gustavo E. Fischman
California State Univeristy–Los Angeles

Richard Garlikov
Birmingham, Alabama

Thomas F. Green
Syracuse University

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Ontario Institute of Technology

Patricia Fey Jarvis
Seattle, Washington

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Les McLean
University of Toronto

Heinrich Mintrop
University of California, Los Angeles

Michele Moses
Arizona State University

Gary Orfield
Harvard University

Anthony G. Rud Jr.
Purdue University

Jay Paredes Scribner
University of Missouri

Michael Scriven
University of Auckland

Lorrie A. Shepard
University of Colorado, Boulder

Robert E. Stake
University of Illinois—UC

Kevin Welner
University of Colorado, Boulder

Terrence G. Wiley
Arizona State University

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 334 times since July 13, 2003

Volume 11 Number 21 **July 13, 2003** **ISSN 1068-2341**

Educational Policy Formation in Loosely Coupled Systems: Some Salient Features of Guatemala's Public and Private School Sectors

Carlos R. Ruano
El Bosque University
Bogota, Colombia

Citation: Ruano, C. R. (July 13, 2003). Educational policy formation in loosely coupled systems: Some salient features of Guatemala's public and private school sectors. *Education Policy Analysis Archives, 11*(21). Retrieved [date] from http://epaa.asu.edu/epaa/v11n21/.

Abstract

The purpose of this article is to analyze the formulation and implementation of educational policy processes in relation to private schools in Guatemala. Specifically, how bilingual education is defined and implemented in the private education sector in Guatemala City where the largest number of privately run establishments exist. Given the great deficits in the provision of educational coverage in the public sector, there has been an explosive expansion of private institutions which have very different levels of quality. Through an analysis of the administrative processes within the Guatemalan Government in general and its Education Ministry in particular as well as of the governance arrangements existing in the private school sector, an overall view of the curricular and policy decisions taken by private schools in the formulation and implementation of bilingual education is presented. This study was based on a sample of six private schools which cater to higher income segments of Guatemala City's student population. Some of the relevant findings of this study include, the existence of a situation of quasi autonomous institutional functioning of the private sector, extreme differentials in the quality of services provided, inadequate levels of teacher and school administrator's training as well as lack of cooperation between public and private sector schools.

Introduction

Many researchers of Education are familiar with Guatemala's multilingual and multicultural traits. That is to say, the presence of large Maya and Spanish speaking populations plus the smaller Garífuna and Xinca linguistic groups. Much less is known, however, about the internal dynamics of Guatemala's educational system particularly in relation to the formulation and implementation of bilingual education policy in privately run schools. In a country of some 10 million inhabitants and a landmass about the size of Switzerland, there are some 22 indigenous languages spoken by half of the population. Nevertheless, this rich cultural diversity is strikingly absent from the school curricula in the private schools (England 1998, Artiles 1995). One goal of this article is to investigate the process of bilingual curriculum formulation and implementation in order to understand the sociolinguistic choices made at the school level. Another goal is to analyze the internal governance processes within the country's education apparatus and their relation to the salient bilingual education curricular and linguistic arrangements deployed by those schools which are privately run.

To accomplish these goals, this article is divided in five sections. First, a sociopolitical overview of the conditions where the education system evolved is given. Secondly, a review of the legislative process within the Guatemalan State is presented including the general legislative framework in relation to educational policy. This refers to the interactions between the main lawmaking body, the National Congress and the Presidential or Executive level of decision making. The third point consists of a description of the Guatemalan government's administrative mechanisms. Such description is given in order to understand the flow of decision making processes within the bureaucratic structures.

This segment also addresses the role played by the Ministry of Education or MINEDUC from an administrative standpoint in the shaping and execution of educational policies and the differential outcomes that Ministerial decisions have for private education institutions. The fourth aspect of this research deals with research design issues such as methodology, fieldwork conditions and the establishments selected for this study. The fifth point is an analysis of the salient characteristics of private bilingual education institutions that cater to higher income population in Guatemala City including their governance, financial, curricular and parental involvement aspects. Lastly, the appropriate conclusions are presented.

For purposes of this analysis, private institutions of education (known as *Colegios Privados* in Guatemala) are defined as those which are for the most part organizationally and financially self-sustaining. Furthermore, the term bilingual education is used to denote schools in which most of the learning activities are conducted in two languages. To this author's knowledge, not a single private school which caters to the middle and higher income segments of the population in Guatemala City defines bilingualism as the inclusion of Maya languages in its curriculum alongside Spanish. For these establishments, bilingual education is regarded as the teaching of Spanish in addition to another European language, generally English.

Private schools located outside of Guatemala City were not analyzed as part of this research due to the fact that the overwhelming majority of establishments of this kind are located there: Guatemala City concentrates over 80% of all private schools in the country (Revista Cronica 1997, Revista Proceso 1998). Lastly, the analysis is centered around schools that cater to the Primary and Secondary Levels. The latter is further divided into Lower Secondary and Higher Secondary or Vocational Track students.

That is to say, those between 7 and 17 years old. The majority of enrolments take place within this age band. While some initiatives are being implemented in terms of bilingual education in Indigenous Maya languages particularly in rural areas, (Enge & Chesterfield, 1996), such initiatives are still in their early stages and cannot be compared with the practices reviewed here.

Sociopolitical overview

> We live in a blind, repressive society with brutal coercion and instinctive passion rather than reason as its guiding principles. In our social environment injustice and lack of respect are the norm while a Neanderthal contempt for ideas dominates us all. (Note 1)
>
> You needn't kill everyone to complete the job...[During the 1980's] We instituted Civil Affairs which provides development for 70% of the population while we kill 30%. (Note 2)

Politically, Guatemala, is organized on a republican system with an Executive or Presidential branch, the Legislative or Congressional arm and the judiciary. Both Executive and Legislative branches are elected simultaneously for a period of four years. By law, the Executive is forbidden from seeking re-election.

Congress is made up of some sixty deputies from all 22 Departments who are also elected to a four-year term. There are no term limitations for deputies. As the legislative branch, Congress either passes laws proposed by its members or by the President who can veto any laws passed. With the exception of the short-lived Serrano government (1990-1993), there has never been a congressional majority made up of deputies from a party different than that which controls the Presidency. Therefore, a Presidential veto is never reversed. This form of government with different political parties alternating in the exercise of power via the electoral process is relatively new. Nevertheless, as recently as 1993, the inability to govern from a divided powers perspective was illustrated during J. Serrano's Presidency , when antagonism between Congress and the Executive resulted in an attempt by the Executive to abolish Congress and to impose rule by decree, Caudillo style. The attempt to shut down Congress failed and Serrano was forced to flee the country.

For the most part of its existence as an independent State, the country has been ruled by a series of strong men or Caudillos. The Caudillo is a persistent feature of Guatemala's sociopolitical landscape. Since Independence from Spain in 1821, with brief respites of democratically elected governments, Caudillos have been the standard feature of Guatemala's Presidential system of government. between 1837 and 1944, four Caudillos ruled the country during 76 years:

José Rafael Carrera 1837-1865
Justo Rufino Barrios 1871-1885
Manuel Estrada Cabrera 1898-1920 and

Jorge Ubico 1931-1944.

Between 1944 and 1954 two democratically elected governments headed by Juan José Arévalo and Jacobo Arbenz respectively, introduced fundamental changes at many levels of society. Expropriation of lands belonging to the United Fruit Company put the governments of Guatemala and the United States on a collision course. In 1954, an invasion force organized by the Central Intelligence Agency entered the country. The Army refused to fight it off and soon thereafter, President Arbenz went into exile never to return (Immerman 1982, Gleisejesses 1991). The subsequent counterrevolutionary regime proceeded to roll back many of the reformist measures of the previous decade. Civil and political freedoms were repressed also. Between 1954 and 1985 all but one of the governments were headed by military men. These governments created and maintained a vacuum in the political center by killing or forcing into exile, leaders not only of left-wing groups and guerrilla sympathizers but of moderate center-right parties, university professors, rural and urban labor union organizers, teachers, business leaders, healthcare workers, artists, intellectuals and Clergy along with anyone else who might be perceived as an emerging civilian leader (Aguilera Peralta 1980, Albizures 1980, Chomsky 1991, Frundt 1987, Levenson 1989).

From 1960 onwards, as leftist Guerrillas began to mount a series of armed operations, the levels of government-sponsored terrorism rose dramatically, culminating in the scorched earth policies of the period 1980-1984 (Falla 1994, Handy, 1992, Lebot, 1992,). Numerous teachers were murdered as part of this policy of extermination. The Army and paramilitary groups targeted them as potential leaders or guerrilla sympathizers (CEH, 1999). By 1990, the entire country was under military control with military detachments and bases in 20 of the country's 22 Departments (Smith,1990). At the height of the carnage, some 25 percent of the country's population was displaced and tens of thousands were killed or disappeared. In areas where the policy of extermination of civilian populations reached its logical conclusion, up to 80 percent of the population was displaced. Ninety percent of the Human Rights violations documented between 1960 and 1996 are directly imputable to the Guatemalan State and its agents (army, police and paramilitary Death Squads). Some 7% were committed by leftist Guerrillas while the remainder 3% cannot be attributed to either party (REMHI, 1998).

Despite relatively stable levels of economic growth during the period 1950-1980, no long term re-investment policies in education or health were developed. Thus, socioeconomic prosperity remained confined to a small segment of the population (Barry & Preusch 1986, CEPAL 1984, Demyck 1983). As early as 1978, a World Bank study had identified the single most important factor in Guatemala's weak performance in several key economic and social indicators: The lack of a well developed education system.

For example [...] Korea's industrial investment during the period 1965-1973 was only three times that of Guatemala, but its industry provided employment to 965,000 people, fifteen times more than did Guatemalan industry.

Critical to the success of Korea's development strategy was the highly developed education system which produced a literate population able to acquire industrial skills quickly. Guatemala's education system needs to be upgraded markedly if the country is to reduce unemployment through the development of industry. In these fields Guatemala's efforts are still far from adequate (p. 16).

As Ibarra de Calix (1997) found twenty years later, the educational system continues to be the Achiles Tendon in all efforts to modernize the national economy:

> Two of the factors which greatly reduce Guatemala's workforce competitiveness are the lack of skilled workers and the low level of training among the general population. [...] Should these trends continue in terms of quality and overall training levels, the country will not be able to overcome its underdevelopment thereby preempting society as a whole from benefitting of the technological advances and transfer of new technologies produced elsewhere. (p. 318). (Note 3)

The resulting cycle of underemployment and under educational achievement has exacerbated the levels of socioeconomic inequalities: Guatemala's Income Distribution disparities are the greatest in the Western Hemisphere second only to Brazil's (CEPAL, 1997).

The legislative framework

> All citizens have a right and an obligation to receive education at the Pre-Primary, Primary, Lower and Higher Secondary levels within the age limits established by law. (Note 4)

> Some laws are made by God and those are untouchable. Some laws are made by men and those can be argued about. Then there are laws made for Guatemalans. Them laws are like hot cinders; good to keep the rich man's house warm but never enough to bring light to a poor man's ranch. (Note 5)

The passing of legislation is a lengthy and convoluted process with lots of procedural delays and wrangling over the most minute drafting technicalities, the formal elements of law clearly having precedence over their actual significance. Hence, only those legislative items which top the presidential agenda receive appropriate attention and are dealt with within a reasonable time frame. Thus, it can take several months or even years for those legislative items not pushed by the President to be passed.

Educational regulations must undergo several layers of legislative approvals before coming into force. The first such layer is Congressional approval. According to Guatemala's legal system, all administrative changes and policies enacted by government departments must be approved by Congress. Such approval usually takes between two or three years depending on the priority given to the changes by the Executive. Because of these lengthy procedural delays, Ministries and sometimes the quasi-ministerial agencies issue binding regulations known as Acuerdos Ministeriales [Ministerial Decrees]. The President can also issue binding regulations; in this case they are known as [Acuerdos Gubernativos [Executive Decrees]. All these types of regulatory acts have the same legal force as laws passed by Congress. In addition, all of these different types of Decrees are issued with or without Congressional approval nor oversight thereof. Such parceling of public policy results in a panoply of intricate and oftentimes contradictory sets of regulations set up by different levels of government

Since the restoration of democracy, most governments have been headed by members of the same political parties in both the Executive and Legislative branches. As a result, all major educational policy initiatives are initiated by the Executive. The inability to govern from a divided powers perspective was illustrated during J. Serrano's Presidency in 1993, when antagonism between Congress and the Executive resulted in an attempt by the Executive to abolish Congress and to impose rule by decree, Caudillo style. The attempt to shut down Congress failed and Serrano was forced to flee the country. In short, educational policies are generally imposed on the educational system top-down fashion with little or no input from the affected parties. Once new legislation is approved, Congress must then issue specific regulations which spell out in great detail the scope and limitations of the acts that can be performed under the new laws. Again, the passing of the specific regulations can take a long time unless it is high on the Executive's agenda. As a result, although Guatemala had several major educational reform initiatives over the last two decades, the actual functioning of the Ministry of Education is still governed by the regulations dating back to 1977. Therefore, administrative and policy changes required to bring MINED's internal organization in line with the legislative changes passed by Congress never took place (Galo de Lara, 1997).

To compensate for the absence of appropriate regulations, a full panoply of ad-hoc Ministerial Decrees has been enacted over the years. In addition, numerous Executive Decrees have also been issued in an equally haphazard fashion and with no apparent policy direction nor long term objectives.

Both of these deal with every conceivable action such as setting up of new interministerial agencies, execution of educational reforms, licences to operate private schools and even to day-to-day administrative matters. Their exact number is unknown. As of 1998 , there was no centralized legal database or catalog of existing laws or newly approved ones.

It was estimated that some 15 thousand pieces of legislation and other 30 thousand Decrees from different levels were neither registered in a database nor properly catalogued (Larra, 1998). These legal entanglements are partly responsible for the considerable delays and institutional weaknesses noted in many aspects of the internal operations at MINED.

Administrative processes in Guatemala's government

Administratively, the country is divided in 22 Departments and some 325 municipalities. Mayors are the only locally elected authorities. They are elected to periods of two to four years depending on the size of their municipality. In addition, there are Departmental Governors who are directly appointed by the Executive. Neither municipal authorities nor Governors have any decision-making input in the educational policy process at the national level. Rural areas participation in national policy making processes is further constrained by the legal and administrative operational definitions which date back to 1938 and remain unchanged (United Nations, 1999).

The national government is organized in a myriad of administrative units of varying size and competencies, oftentimes with high degrees of duplication. At the top there are the Ministers who are appointed by the Executive. These individuals are almost always appointed to their positions as a result of their loyalty to the President rather than by their professional suitability for the post. Underneath the ministerial echelons there are Deputy Ministers [Vice-Ministros in Spanish]. They are in charge of day to day operations of their departments. Again, loyalty to the person who appointed them rather than competency or the fulfilment of organizational priorities is the main criterion for appointment to the position. This is a feature of public administration in Pre-Modern States known as the Loyalty Principle (Ruano, 1999). Simply stated, the Loyalty Principle posits that regular bureaucratic channels which support organizational control mechanisms, lines of accountability and policy implementation are bypassed in favor of decision making processes based almost entirely in the pre-eminence of personal ties to the individual from whom appointees derive their power base. In this type of Pre-Modern administrative arrangement, people are not expected to work for the objectives of the bureaucracy but for the person who appointed them (Ruano 1999 p. 2ff).

A third level of administration is composed by the Directors General who oversee specific agencies within a Ministry. Each Ministry has many agencies which operate in a quasi-autonomous manner and are known as Direcciones Generales [General Directorates].

These Directorates and Sub-Directorates have branches throughout the nation where further atomization of public policy has been identified by previous research (Dignard 1987, Galvez-Borrel 1996, Dunkerley 1988,). Other administrative shortcomings observed include critical shortages of qualified managers, high turnover, excessive duplication, inadequate taxation rates and feeble fiscal accountability procedures, lack of clear hiring and promotion criteria, nepotism, constant shifts in administrative and public policy priorities, weak enforcement capabilities and corruption (Brewer-Carias 1979, Clark 2000, El Periodico 1998, Heyman 1995, Handy 1991, International Monetary Fund 1995, United Nations 1998). Other government agencies with quasi-ministerial characteristics include the Taxation Administration Authority, the Judiciary, National Housing Administration and dozens of others.

All ministries follow a similar organizational pattern except the Ministry of Defense which has its own internal arrangements and for all practical purposes is only accountable to the military High Command (Black 1985, Goldman 1999, McClintock 1985, Nairns & Simon 1986 Ruano 1997). (Note 6) Furthermore, within this organizational pattern, numerous instances of Loose Coupling (Churchill et al.1979, Gamoran & Dreben 1986) are observed. That is to say, patterns of administrative behavior whereby managers and employees are mostly concerned with the operational survival of the units under their control rather than the overall functioning of the Civil Service. This results in further diffusion of lines of accountability at all levels of the government. The specific consequences of these regulatory arrangements for the implementation of bilingual curricula in private schools are also analyzed in this research.

MINED's structure and role in the education system

> The defining characteristics of the education sector are inequity, low coverage and low quality (Note 7) .

In principle, the Ministry of Education or MINED is responsible for educational policy, including planning and coordination as well as curriculum design and quality assurance for both public and private education. MINED supervises all levels of instruction [kindergarten, primary, lower and higher secondary] except Higher Education which is self governing due to Constitutional Mandates which grant Self-Government or *Autonomia Universitaria* to the University. MINED is Guatemala's largest government department in terms of its number of employees, overtaking Defense and Public Health Ministries. Following the organizational pattern found elsewhere in the Guatemalan government, MINED is composed of a myriad of Departments, General Directorates and autonomous units all of them functioning with high degrees of loose-coupling. According to a World Bank study (1995), some 1,400 agencies, units or departments were found to be under the nominal control of MINED. Many of these units had been created to oversee specific projects or multilateral agreements and continued to exist long after their original purpose had ceased to exist. Between 1980 and 1990, the number of teachers remained unchanged while administrative personnel increased by 15 percent (UNESCO, 1991).

At the same time, other Ministries such as Agriculture, Health and Defense maintain educational facilities of their own and, in the case of Defense, they are in charge of training for large numbers of conscripted soldiers in non-military occupations inside military installations. All these expenditures and resources deployed are outside the control of MINED or its oversight.

There is also a large number of Spanish language schools which cater to foreigners. These schools are under the control of the Guatemalan Institute of Tourism which issues licences authorizing the functioning of such facilities.

There is a sharp division of tasks between MINED's administrative employees and teachers. The latter have little or no input in any administrative and procedural matters nor is their input sought when drafting educational reform initiatives which are always prepared by administrative personnel with the assistance of international agencies or consultants.

Moreover, duplication and compartmentalization of the simplest tasks can reach extraordinary proportions, forcing even the simplest administrative decisions to go through a number of steps and procedures. For instance, it takes up to two years after graduation for a student to receive her post-high school or vocational track graduation diploma, such is the number of signatures and approvals required. As no clear lines of authority are defined, and given the patron-client relations established between different individuals within MINED, diffuse decision-making is constantly exercised. In practice this means that nobody can or wants to assume responsibility for anything that is not clearly outlined in a legal procedure. Simple decisions take months to be made -if at all. Many of these individuals report directly to the Minister who is perceived as the final decision-maker. This perception is not groundless; he or someone acting on his authority is required to sign and approve every single appointment and promotion at all levels. These micro management traits result in lower level employees fearing for reprisals from their superiors should the approval of the latter not be obtained before making a decision however simple it may be. These restrictive practices result on reduced level of institutional accountability while increasing the isolation between MINED's agencies. Passive resistance becomes the behavior of choice for those whose livelihood depends on the existence of their unit in isolation from all the others. As one MINED employee put it, "I can feed the Minister three kinds of information: false information, misleading information and the truth. None can force me to give him what he wants." All these characteristics tend to reinforce the public's perception of MINED as a non-responsive, closed institution with little or no regard for their concerns.

Hierarchically speaking, MINED is essentially a top-down structure consisting of four strata. At the top level we find the Minister and about ten or so associates who are recruited by him on the basis of loyalty and -only secondarily- ability. Some of these individuals are not classified as regular civil service personnel. Instead, they are known as Advisors (*Asesores*) to the Minister while others are appointed as technical and administrative Deputy Ministers.

This group advises and shapes the general orientation of the Ministry in accordance with the Minister's priorities. Those at the top level regard all levels below them as unmanageable dead weight "unless you work with a group of like-minded technocrats or bring in your own people from outside, it is very hard to get anything done at MINED" (Galo de Lara, personal communication). The interests of the top echelon are fully political, tied as they are to the fortunes of the Minister who brought them on board. Their time in office is unpredictable, ranging from a few months to four years as all new Ministers bring in their own group of advisors and no government carries on the educational policies of previous ones.

Below the Minister and Deputy-Minister levels, all Departments are staffed by civil servants. The third level is occupied by the Directors General at the national level. As a result of administrative regionalization in the mid-1980's, further departmental directorships were created. There are some twenty units headed by a Director General. The units in charge of primary and secondary/vocational education are the *Dirección general de educación primaria* and the *Dirección general de educación media* respectively.

The fourth level in the administrative chain is the Supervisor. They are the only operational link between private schools and the MINED bureaucracy. In 1999, there were some 55 supervisors in charge of overseeing all of the elementary and secondary schools in Guatemala City. (Note 8) The figure given in the text was compiled by the author from interviews during fieldwork throughout Guatemala City.. Their main task is to enforce ministerial policies and regulations at the school level for they constitute the only direct link between MINED and school principals in the private and public schools. Each supervisor is in charge of a school district (*distrito escolar*) which is simply a group of schools treated as a unit due to geographical proximity. There are approximately 93 districts in Guatemala City. The number of schools within each district varies widely depending on its size, some covering 30 schools while others covering 50 or more. Supervisors are usually former school teachers with no additional pedagogical nor administrative training and whose oversight functions entail large amounts of time spent on revising paperwork submitted by the schools under their charge. Those individuals who are promoted to the Supervisor's position receive no salary increases nor any other type of incentives to further their professionalization. Over the last three decades few if any training programs or mechanisms specifically aimed at improving the long-term performance and/or qualifications of Supervisors' have been set up by MINED. MINED does not allocate any additional resources to carry out the supervisor's duties. As a result, in many areas of the country, it is common for private schools to contribute with some of his/her expenses, for instance, office space, stationary supplies and clerical support. Several school districts hire secretarial help and donate office space for supervisors. In other instances, such support extended to "salary supplements". In a chronically underfunded Ministry, this is hardly surprising. Even less so, given the regularity of complaints from teachers about late or no payment of their salaries (Siglo XXI, 1999).

Theoretically, the Directorates of Primary and Secondary Education oversee all private schools. In practice, however, not even the Directors General of the Primary and of the Secondary Divisions can agree on a common agenda to harmonize those aspects of the policy process which affect both these divisions even though all supervisors in Guatemala City are required to oversee both primary and secondary schools within their individual districts. Thus, it is not uncommon to find children and teachers without schools; schools unable to have facilities of their own, while empty buildings await those who were supposed to occupy them; empty installations bereft of basic equipment or textbooks and poorly trained personnel; these are some of the consequences of MINED's organizational configuration which were identified by a UNESCO study in 1980. Sixteen years later, another study (Ruano de Flores, 1997) found that the problems identified in 1980 had become even worse. The most recent manifestation of this administrative inconsistency took place during the failed attempt to incorporate students from the senior year in vocational career tracks into the national literacy campaign. The campaign was coordinated not through MINED but through the National Directorate for Literacy and the Presidential Secretariat for Social Affairs, an office traditionally used as an executive branch outlet to give visibility to the President's wife through charity projects. Needless to say, the decision to incorporate the students had been taken without consulting schools, parents nor the students concerned. As word began to spread out that students would be required to give up part or all of their senior year of studies to participate in the literacy campaign, widespread protests erupted throughout Guatemala. In the resulting fiasco, the government was forced to back down and had to redefine the entire scope of the literacy campaign (Prensa Libre, 2001).

Lastly, MINED has very limited research and development capabilities. MINED does not foster any long term cooperation initiatives between public and private schools at either the Departmental or National level. Research comparing curricular, administrative or financial aspects of private and public education in Guatemala is virtually non-existent (Ruano, 2002). There are no formal mechanisms to allow private and public schools to exchange information on best practices, sharing of facilities such as libraries, information technology or teacher professional development programs.

Researching Education in Guatemala: Fieldwork, methodological and sampling issues

The majority of fieldwork was undertaken between 1997 and 1999 as part of a Doctoral Dissertation. Subsequent observations and follow up interviews took place in 2001 through a University Research Grant. Some of the schools observed follow the standard Guatemalan Academic Calendar from January through October while others use the North American one from September through June. While the original fieldwork included observations, interviews and statistical analysis of some 40 private and public schools throughout the City, the findings presented here are based on work carried out in six private establishments which shared similar characteristics as follows:

> Between 700 and 1000 students total enrolment more or less evenly divided between Primary and Secondary Divisions (between 7 and 17 years old).
>
> Fully Bilingual Curriculum from first grade Elementary to Senior year Secondary where English is the Second Language taught and Spanish the primary one.
>
> Total financial autonomy from public funding sources. Fully funded from student fees and other types of private contributions.
>
> Fully self-governing through a Board of Trustees or as a family run and owned institution.
>
> Non denominational.

During fieldwork preliminary open ended questionnaires were administered to both teaching and administrative staff in order to get baseline data on teaching assignments, administrative tasks and student performance indicators. Further data was obtained through focus groups with parents, students and administrative and ancillary staff. Data on Governance arrangements was obtained via interviews with key decision makers in the schools. Documentary cross checking was carried out through examination of schools records, Ministry of Education documentation (when available) and related materials. Nevertheless, the research effort faced clear and at times severe restrictions. Some of the most important involve categorical refusals on the part of most schools to go on record as to financial and governance aspects. Not a single school from which data was obtained in this study would allow its name or the identities of its Officials to be named. Only exceptionally was the researcher allowed to make copies of school records. With rare exceptions no interviews were taped and no participants in focus groups or individual interviews could be named or implied if such implication could lead to their identification. Other restrictions involved agreements made between the researcher and the schools so as to prevent sharing of Curricular and pedagogical practices with other schools.

Some of the arguments given by school officials regarding the restrictions imposed are related to security concerns; Guatemala City is regarded as one of the most violent regions in the world (World Bank 1997, Buvinic, Morrison, & Shifter 1999) sharing top murder and kidnaping rankings in a list where one can find cities such as Bogota (Colombia), San Salvador (El Salvador) and Johannesburg (South Africa). Evidently, some of the students enrolled in these schools are prime targets for extortion and kidnapping and their security becomes a basic concern at the schools they attend. Other rationales given are related to economic considerations, as in the case of restrictions imposed on the sharing of Curricular and pedagogical practices with other schools. Such practices are regarded as "trade secrets" by the schools. Hence their reluctance to have those practices known by other schools regarded as competitors in a market segment that is very small due to the extreme socioeconomic stratification noted previously. Based on these limitations, the data and analysis thereof is presented in a composite-type fashion from which an overall picture emerges. That is to say, the data is integrated into a general framework from which analytical categories can be discerned and conclusions drawn. For instance, rather than focusing on specific financial or governance arrangements of individual schools, an overall analysis is presented.

Lastly, where it is necessary to refer to the larger context of private education in Guatemala, the analysis presented is based on the baseline data obtained throughout fieldwork. Due to the scarcity of previous research in this area, much of the data was generated and is presented in a systematic fashion for the first time.

Private bilingual education: organizational and governance aspects

> Look, what do you think this is? An American Indian Reservation or something? Our students' parents want their kids to learn English, not Mayan languages. (Note 9)

In terms of ownership arrangements, elite bilingual private schools in Guatemala City can be divided in three large categories. First there are those who are run as family businesses by members of a family. In this model, certain administrative tasks are delegated to a Principal or School Coordinator while the owners retain overall control in all other aspects including, textbook selection, hiring and firing of teachers, performance assessments, academic and other fees charged to students as well overall curricular orientation. Members of the owning family usually hold positions such as teachers, counselors, librarians or accountants for the school.

The second category is made up of establishments that are run by a Board of Governors (Consejo Directivo). This board is usually made up of parents and other individuals who have contributed financially to the setting up of the school and can in fact be construed as shareholders with overall control of the school's activities. Usually, the board hires a Principal and other administrative staff upon whom many of the day to day functions are delegated.

The third category is made up of those schools which are controlled by religious or denominational entities. Historically, the majority of denominational schools are from Evangelical and Catholic denominations (Rose & Brouwer, 1990). In recent years, the Church of Latter Day Saints or Mormons has set up schooling facilities for its members throughout Guatemala. In the case of Church owned establishments, overall control is retained by the religious entity with limited or no input from non-Church participants. All of the schools where data collection took place for this study where non-denominational. As a general rule, denominational schools are quite difficult to gain access to due to a number of factors such as mistrust towards outside researchers, lack of interest in educational practices not approved by their superiors and an overall climate of defensiveness on the part of school officials. A very small segment of the privately run schools is made up of so-called bi-national schools. These were set up by German, French and United States immigrants and expatriates resident in Guatemala.

The binational schools tend to follow the curricular arrangements of the countries they are associated with, granting exit diplomas comparable to those of schools in the home country. Their clientele is made up of wealthy Guatemalan families children of expatriate workers and descendants of immigrants.

In terms of organizational arrangements there is also variability. While the establishments dealt with in this study have a multi-layered organizational pattern composed of teachers, principals and administrative support staff, the vast majority of private schools have a simple structure comprised of teachers and a principal who can also be the school's owner. Affordable private education usually consists "of large numbers of students crammed into matchbox-sized classrooms with school facilities and teachers not worthy of the name" (Fernández García, 1998, p. 4). A standard private school is usually a reconditioned home with a small courtyard and few -if any- pedagogical aids. It must be remembered that, with limited public investment in education in previous decades, demographic pressure forced many lower-income families to enrol their children in these private institutions which were the only ones they could afford and are still preferable -in their view- to many public schools (Galindo, Personal communication).

Supervision of private schools takes place on paper only. This means that they are required to present large amounts of forms and documentation attesting to the schools' program content, teaching staff and facilities. Nevertheless, supervisors will seldom visit a school to verify the validity of the claims made. Once a school is registered its licence to operate is granted through a Ministerial Decree in some instances while other schools are authorized under an Executive Decree. It is unclear why some establishments were authorized to operate under different types of Decrees. Nevertheless, such licenses must be re-validated on an annual basis.

This means that Supervisors spend a great amount of time dealing with the paperwork generated by private schools when renewing their licenses.

Curricular organization and teaching aspects

In terms of Curriculum formulation, the Education Law requires a threshold of compliance with the basic curricula which is comprised of Mathematics, Spanish Language, Social Studies and Biology whenever they are taught in Spanish. Any other activities, subjects and programs of study delivered beyond the basic curricula and delivered in languages other than Spanish are considered to be optional and are neither subject to administrative nor academic oversight. Overall compliance with the basic curricula is enforced by MINED supervisors. As has been noted earlier in this section, Supervisors have very limited enforcement capabilities to bear upon the very institutions which oftentimes supply them with basic office equipment and "salary supplements". As a result there is virtually no way of enforcing compliance with basic curricular content. Furthermore, private schools are at complete liberty to set up their own admissions standards which can include entrance examinations, financial screening and personal interviews with both parents and prospective students. Results of examinations are never made public and there is no mechanism to compare performance across schools. There are no nation-wide standardized tests, process indicators nor any other recognized input or output indicators used with enough consistency to allow for proper assessments of the private education sector to be carried out. This is an institutional weakness which is present in various degrees throughout Latin America (Birdsall & Sabot 1996, Dignard 1986, Lowden 1996, Otis 1997, Savedoff 1998, Silva 1996, Vos 1996).

Generally speaking, Private institutions have an almost unlimited discretion in curriculum design and teachable subjects so long as schools claim that part or all of their instruction takes place in a foreign language. For example, almost all private schools claim to offer one or more bilingual -mostly English/Spanish- vocational tracks. Because English is considered to be an additional subject it is usually off limits to ministerial oversight.

Thus, most private schools' advertisements profess to be "fully bilingual" meaning English-Spanish. In relation to the establishments analyzed for this study, the preferred approach is to set up a situation of parallel bilingualism whereby a certain number of subjects is taught entirely in English alongside their Spanish counterparts. The main curricular and governance consequence of such arrangement is that a school is effectively split in two separate areas along linguistic lines. Thus, there is a Principal in charge of the English language segment while another runs the Spanish one. Little or no coordination was observed to exist between both segments.

Hiring and firing of teachers takes place at the discretion of the school's owner with no right to appeal by the teacher. None of the schools researched by the author throughout Guatemala City have seniority nor tenure provisions for their teaching staff. Thus, teaching salaries remain stagnant or accrue only small increases over the span of many years of service. Retirement plans or Pension Funds for private school teachers are virtually unheard of. There are no legal avenues to counter these arbitrary practices as the working conditions and contracts set forth by private schools are not regulated by MINED but by the Ministry of Labor. There is no professional organization to represent private school teachers nationwide and no initiatives to improve their professional standing and working conditions have ever been undertaken by MINED or other stakeholders.

Working conditions for most teachers are difficult. As salaries are low, it is not unusual to find teachers who work in two or even three different jobs to make ends meet. This is a critical factor behind low teaching standards and poor student performance. It is also a phenomenon found in many developing countries (Farrell & Oliveira, 1993). Even the elite bilingual schools studied spend little or no resources for professional development. Not a single private school visited during fieldwork offers financial incentives or time off support for teachers who wish to further their training at the university level. Similarly to their public sector counterparts, private school teachers need no additional training nor credentials to become principals at either the Primary or Secondary level (AMEU 1998, Fadul 1997). Likewise, pay scales, in-service training and additional support mechanisms for teachers are left entirely at the discretion of the establishment's owner whose ultimate decision-making authority is never questioned. As pointed out previously, school owners are rarely qualified school administrators. "My father gave me my *Colegio* as a wedding present" said the owner of a school interviewed by the author. During fieldwork, this author observed how principals almost invariably deferred to the establishments' owners opinion even in areas which were clearly Curricular in nature.

Financing the autonomy of private schools

Due to constantly increasing demand for educational services, private schools became an important provider of educational facilities. According to article 73 of the Constitution, all private educational establishments are exempt from payment of all taxes on all their activities. This exemption extends to school supplies and materials, tuition fees and ancillary fees as well as to infrastructure. In the case of schools looked at in this study the range of fees varies between three and five-thousand dollars a year. Thus, total tax exemptions amount to millions of dollars. It is important to point out that private schools receive no direct funding from the government. Therefore, each of them must meet its financial needs through registration and ancillary fees charged directly to students. There is no publicly funded system to support attendance of children from lower income background whose parents wish them to attend private schools

Demographic pressures during the 1970's and 1980's coupled with declining government investment in the construction of new schools created an ever increasing demand for privately-run facilities. The growth in private education has not been matched by national regulations on the quality and type of instruction offered. As a result, schools tend to focus more on their economic viability rather than their academic performance. This is true of the vast majority of establishments and probably more so of the elite bilingual schools whose only claim to superior academic standards lies in the prestige levels they project to society. In other words, brand name recognition reinforced by perceptions of economic success of the families of attending students have more weight in a decision to choose a school over their competitors. Many parents in focus groups were particularly explicit on their motivations in selecting individual establishments. In the families researched in this study, such selection is usually made by the mother as opposed to the father. Among the rationales given one can find:

> "That's where I went to school, therefore, that's where my children will go."
>
> "I heard from other mothers that this school has a good reputation."
>
> "I know the Principal of this school. She's very good."
>
> "The school is in the neighborhood. You know, one less security risk to be worried about."
>
> "I want my children to learn good English and math. This school is one of the best in preparing kids to go to universities in the United States and succeeding in those academic environments."
>
> "Most of the families of my children's friends have sent their kids here. It's a way of keeping things in the family so to speak."

To a certain extent, attendance at these elite establishments reinforces and reproduces the social networks which are necessary to maintain one's own position in the social hierarchy. In a social environment such as Guatemala's where socioeconomic advancement prospects are conditioned by the Loyalty Principle rather than by individual talent, educational criteria take a back seat to socioeconomic status considerations. It is not surprising then that, with some exceptions, most parents did not seem to be interested in having nationwide educational standards that could allow them to compare schools based on clear performance indicators.

"We know just which schools would come on top of any classification anyhow. So, no, I don't see a need to classify them" said a confident upper class parent when asked about educational standards. While it can be argued that subsidizing private education via tax exemptions is necessary and even desirable in light of the large coverage and quality deficits extant throughout the country, it can also be argued that some or all of these deficits would not have arisen had the government been able to generate enough revenue via general taxation revenues to finance the quantity and quality of additional public schools. Guatemala's overall tax collection rates are among the lowest in the Americas (World Bank, 1995). Moreover, a great deal of resources was spent for military purposes particularly during the period 1980-1995. The resulting diversion of resources away from educational activities have turned MINED into a chronically underfunded Ministry with no hope of reversing this trend in the foreseeable future. Five years after the end of Civil War, Military expenditure still consumes substantial portions of the country's budget.

In addition to tax breaks, private schools obtain substantial revenues from several activities. The most widely used are:

- Entrance examinations (*Exámenes de admisión*)
- School supplies lists (*Listas de útiles escolares*)
- Registration bonuses
- Fundraising events.

These sources of income will be discussed in turn.

Entrance examinations. The majority of private schools conduct entrance examinations at the beginning of the school year. These tests are not standardized or regulated by MINED in any way. As is the case in most private school decisions, the criteria used for grading these tests are not made public nor are their results subject to appeal. Essentially, each school decides what it is going to test and how much students must pay to sit the examinations. In practice, entrance examinations fees operate as a preliminary financial screening device for parents: if they cannot afford to pay the entrance examinations fees to begin with, then their children are not suitable for attendance at that school. Usually, the higher the cost of school fees, the higher the cost of the entrance examinations. The examination itself is used as a negative selection device in that it reflects the school's own curricular standards thereby preempting access by students who were not exposed to such prior knowledge.

School supplies lists. At the beginning of the school year, each private school decides which school supplies are to be purchased by the students. Most of those supplies can only be found at the school store or from suppliers approved by it. Each student is expected to follow the lists' requirements to the letter. Lists are outrageously detailed and are viewed by both parents and teachers as an outward sign of the school's academic standard. In most cases, schools buy new textbooks and other materials and lease them to students during the school year. The same materials are leased for several years. Through such leasing schemes, the school recuperates the costs of materials several times over. Periodically, MINED threatens to impose fines and other sanctions on schools whose lists are deemed utterly extravagant in pedagogical and financial terms. Nevertheless, it appears that no such action has ever been taken.

Registration bonuses. Registration bonuses are set amounts which parents are required to pay upon first registration of their children. A number of justifications are given for this charge, among them, infrastructure expansion, decreased enrollments, increased enrollments or some other unexpected contingency. It is worth noting that teachers' salary increases is not among the reasons for requesting these bonuses. The amount payable appears to be completely arbitrary and to depend exclusively on the schools' authorities. One parent described the bonus as a "dowry payment we make so that the school will find my child acceptable, so to speak." Under the current legal framework parents have very limited ability to challenge decisions made by private schools.

Fundraising events. Fund-raising events are activities organized by schools to generate income for a variety of reasons. Again, they are decided upon by the school with little or no input from parents. Many schools simply ask of students a certain amount for "Fundraising activities" for the school year, in addition to the bonuses discussed previously. Since few if any private schools release yearly financial reports, it is impossible to know the actual destination of these and any other "contributions" demanded from students.

Concluding remarks

Research to understand the differential impact of educational policy and curriculum formulation in private institutions in Guatemala is at best very limited. There are neither formal nor institutional mechanisms to allow private and public schools to exchange information on best practices, sharing of facilities such as libraries, information technology or staff development programs. No initiatives to promote cooperation between the two sectors have been implemented. Despite their stated goal of focusing on quality and the improvement of academic standards, no comparative assessments between private schools exist. Thus, parents have no way of comparing test scores, repetitions and dropout rates or costs per student across establishments. Rodolfo Bianchi, former President of the Association of Private Schools could not offer an explanation as to why such comparisons were unavailable (Personal communication).

In the absence of any common guidelines for quality assessment, national standardized testing procedures and teaching performance indicators, ability to pay becomes the only criteria separating the different types of private schools. A school's perceived quality is thus inextricably linked to its costs. Ability to pay as the sole criterion to measure for quality also results in academic tracking systems which mirror those of society at large. Only those children from higher income families can afford to attend the same schools attended by their parents. The rest of the student population must rely on the assurances of each individual school as to its educational standards.

Bilingualism in the elite schools is regarded by all parties concerned as primarily the attainment of English language skills at a level which meets parental expectations of social mobility or the preservation of social standing. Other languages, particularly Indigenous ones, need not apply. There is no discussion of the role played by Maya, Xinca or Garifuna languages in the educational reality of the country (Ruano 2001, AVANCSO 1998). Thus, Guatemalan elite bilingual schools tend to reflect the different realities Guatemalans are forced to live in. On the one hand a world of entrenched privilege and dependency on foreign markets which make the survival of exclusionary sociopolitical arrangements possible. On the other, a world of great destitution and inequality which is also endowed with great cultural diversity and resilience.

In terms of financial arrangements, private schools are exempt from all taxation duties. Furthermore, creative accounting practices and other devices insure that financial records remain virtually immune to government scrutiny. At the same time, no private institution in Guatemala City receives direct funding from the government and there is no system to support attendance of children from lower income backgrounds whose parents wish them to attend private schools. In addition, many institutions raise funds through different measures such as obligatory purchases of school uniforms, exorbitant graduation fees, entrance examination payments and ancillary fees. Others purchase textbooks which they then rent out to students for several years at handsome margins. Still others require students to pay hefty entrance fees, the so-called Bonuses (bonos in Spanish). Few if any higher income schools have ever been charged under the provisions of the education law for these abuses (Vásquez, 1998). Though not justifiable, these measures are partly the result of the extremely low priority accorded to education by the Guatemalan government. In the ensuing climate of survival of the financially fittest, private schools have little choice but to make use of such practices.

As for the formulation and implementation of curricula, the lack of clear administrative and legal guidelines allows private schools to decouple themselves from most decision making and policy-formulation processes emanating from the government. For purely ideological reasons, (private schools' perceived conservatism and pro-business stance, attendance by higher social strata, differential fees charged), the Guatemalan government simply assumes that private schools are of better quality than public ones. This insures little scrutiny from both MINED's authorities and the public at large. This perception is also shared by all major international development agencies who usually do not include the private sector initiatives in their overall designs of educational reform packages. Few explanations are offered as to why private schools should neither participate nor be asked for their input in the educational policy process. This exclusion from the policy process is very difficult to understand given that there are five times more private schools than public ones serving the 12-17 age group and that virtually all pre-primary establishments are private (Rodas Martini, 1998).

For almost 40 years, Guatemalans were engaged in a brutal Civil War that essentially originated in the socioeconomic chasm which separates the haves from the have nots. After decades of military stalemate, the conflict ended in 1996, nevertheless, without resolving the outstanding social, economic and political issues which originated it in the first place. With the war over, Guatemalans must now decide whether to pursue social change through the avenues of the democratic electoral process and through increased participation in the national life. In this sense, true and meaningful educational reform is one of the key instruments in the attainment of a more equitable and prosperous outlook for future generations.

Whether Guatemalans are prepared to leave behind the disheartening premises of the present educational system and replace them with alternatives that hold more promise for the future is a question only time can answer.

Notes

1. Press Editorial in *Diario el Gráfico* newspaper (1977) by widely respected journalist and center right politician Jorge Carpio Nicolle. Mr. Carpio Nicolle was murdered in 1993. His assassination bore all the hallmarks of Army Death Squads operations. To this day, the identity of his killers remains unknown.

2. Former Guatemalan Defense Minister General Hector Gramajo quoted in *NACLA Report on the Americas*

(July, 1995). Vol XXV: p.2.

3. Unless otherwise noted, all translations are the author's responsibility.

4. Article 74th of Guatemala's Constitution.

5. Remarks made by a rural teacher in the Eastern Department of Chiquimula on Guatemala's Constitutional provisions which make attendance to school compulsory for all children aged 7-17.

6. The Guatemalan Army's obsession with social control and political repression is still a central element to understand the current socioeconomic outlook. Several years after the end of the Civil War the Army's network of political terrorism headed by the nefarious Military Intelligence Services or G-2, along with several other agencies, remains remarkably intact.

7. World Bank: (1995). Guatemala, basic education strategy: equity and efficiency in education. *Latin America and the Caribbean regional office report No. 13304-GU*. Washington, DC: Author, p. I.

8. Even the exact number of supervisors is not clear. MINED's comptroller' office gave a total of 58, while the Supervisor's office said there were 50 with the rest being retained as auxiliaries or for some unspecified reason.

9. Remarks made by a Vice-Principal during an interview with the author in Guatemala City during fieldwork in 1999.

Personal Communications:

Interviews with Mr. R. Bianchi and Ms. L. Galindo, former President of the Association of Private Schools and School Principal, respectively. Both interviews took place during fieldwork in Guatemala (1998-2001). M.C. Galo de Lara, former Deputy Minister of Education also provided the author with much useful information on the organization of MINED.

References

Albizures, M.A. (1980). Struggles and experiences of the Guatemalan trade-union movement. *Latin American perspectives*. 7, 2-3:145-159.

AMEU [Asociación Maya de estudiantes universitarios] (1998). *La Universidad. La situación de la educación superior en Guatemala en el fin de siglo*. Guatemala City: Author.

Artiles, A.J. (1995). Un reto para Guatemala ante el nuevo milenio; educación para una sociedad multicultural. *La educación*. 121, 2: 213-227.

AVANCSO [Asociación para el advance de las ciencias sociales en Guatemala] (1998). Imágenes homogéneas en un pais de rostros diversos. El sistema educativo formal y la conformación de referentes de identidad nacional entre jóvenes guatemaltecos. *Cuadernos de Investigación*. #11. Guatemala City: Author

Barry, T. & Preusch, D. (1986). *The Central America fact book*. New York: Grove Press.

Black, G. (1985). Under the gun. *NACLA Report on the Americas*. 19, 5:10-25.

Brewer-Carías, A.R. (1979). Les conditionnements politiques de l'administration publique dans les pays de l'Amérique latines. *Revue internationale des sciences administratives*. 3:223-233.

Birdsall N. & Sabot, R. Eds. (1996). *Opportunity foregone. Education in Brazil*. Washington, DC: InterAmerican Development Bank.

Buvinic, M., Morrison, A.& Shifter, M.(1999). Violence in Latin America and the Caribbean: A Framework for Action. *Inter-American Development Bank, Sustainable Development Department*. Internet Document address: http://www.iadb.org/sds/publication/publication_515_e.htm

CEPAL [Comisión Económica para América Latina] (1997). *The equity gap. Latin America, the Caribbean and the Social Summit*. Santiago, Chile: Author

CEPAL (1984). The crisis in Central America: its origins, scope and consequences. *CEPAL Review*. 22:53-80.

Clark, C. (2000). Land Tenure, delegitimation and social mobility in tropical Peten, Guatemala. *Human Organization*. V.59, (4):419-428.

CEH [Comisión para el Esclarecimiento Histórico] (1999). *Guatemala: memoria del silencio. Informe de la Comisión para el Esclarecimiento Histórico*. Guatemala City: Author.

Demyck, N. (1983). Ambitions militaires et esprit oligarchique au Guatemala. *Annales des pays d'Amerique centrale et des caraibes*. 4:7-16.

Dignard, L. (1987). *Administration et développement au Guatemala*. Unpublished manuscript, University of Ottawa.

Dignard, L. (1986). *Education and society in twentieth century Mexico*. Unpublished manuscript, University of Calgary.

Dunkerley, J. (1988). *Power in the isthmus: A political history of modern central america*. London: Verso.

Chomsky, N. (1991). *Pirates and Emperors. International terrorism in the real world*. Revised Edition. Montreal: Black Rose Books.

Churchill, S. Orlikow, L.Greenfield, T.Rideout, B.(1979).*Cost models of bilingual education in Canada. The world of theory*. Final report to the Secretary of State, Canada, contract No.470-410.

El Periodico Newpaper (1998). *Más de 40 familiares de funcionarios laboran en el Estado*. August 10th, pp. 1-3. Guatemala City: Author.

Enge, K.I. & Chesterfield, R. (1996). Bilingual education and student performance in Guatemala. *International journal of educational development*. V.16, (3): 291-302.

England, N. (1998). Mayan efforts towards language preservation. In *Endangered languages*. L.A. Grenoble & L..J. Whaley, Eds. Melbourne: Cambridge University Press, pgs. 99-116.

Fadul, S. (1997). La formación de maestros en Guatemala. En *Reforma educativa en Guatemala*. Guatemala: ASIES, pgs. 213-240.

Falla, R. (1994). *Massacres in the jungle: Ixcan, Guatemala, 1975-1982*. Boulder: Westview Press.

Farrell, J.P. & Olivera, B.J. (1993). Teachers in developing countries: Improving effectiveness and managing costs. *EDI Seminar Series*. Washington, DC: The World Bank.

Frundt, H. (1987). To buy the world a Coke: Implications of trade-union redevelopment in Guatemala. *Latin American perspectives*. 14, 3:381-416.

Galvez Borrel, V. (1996). Guatemala, nueva derecha, viejos problemas. *Nueva Sociedad*. 142:6-11.

Galo de Lara, C.M. (1997). *Diagnostico del sistema educativo en Guatemala* . Guatemala: ASIES Publications.

Gamoran, A, & Dreeben, R. (1986). Coupling and control in educational organizations. *Administrative science quarterly*. 31:612-32.

Gleisejesses, P. (1991). *Shattered hope: the Guatemalan revolution and the United States, 1944-1954*. Princeton: Princeton University Press.

Goldman, F. (1999). Murder comes for the Bishop. Was the most prominent human-rights activist in Guatemala killed by a homosexual lover, a gang of church thieves or an Army hit squad? *The New Yorker Magazine*. March issue, pgs. 60-77.

Handy, J. (1992). Guatemala: A tenacious despotism. *NACLA Report on the Americas*. 26, 3:31-37.

Handy, J. (1991). Anxiety and dread: state and community in modern Guatemala. *Canadian journal of history*. 26, 1:43-65.

Heyman, P.B. (1995). Should Latin American prosecutors be independent of the executive in prosecuting government abuses? *The university of Miami interamerican law review*. 26, 3:535-560.

Ibarra de Calix, I. (1997). *Modalidades de financiamiento de la educación* [Modes of education financing]. Guatemala: ASIES Publications.

Immerman, R.H. (1982). *The CIA in Guatemala: the foreign policy of intervention*. Austin: University of Texas Press.

IMF [International Monetary Fund] (1995). Guatemala: Recent economic developments. *IMF Staff country report, 95/57*. Washington, DC: Author.

Larra, M. (1998). Guatemala sin registro de leyes. *Prensa Libre* Newspaper: July 20th, p. 5.

Lebot, Y. (1992). *La guerre en terre Maya: communaute, violence et modernité au Guatemala, 1970-1992*. Paris: Editiions Karthala.

McClintock, M. (1985). *The American Connection: popular resistance and repression in Guatemala*. London: Sage Press.

Levenson, D. (1989). The murder of an Actor and a Theater. *NACLA Report on the Americas*. Vol. XXIII (3): 5.

Lowden, P. (1996). The escuelas integrales reform program in Venezuela [pp 119-136]. In *Implementing policy innovations in Latin America*. Antonia Silva, Ed. Washington, DC: InterAmerican Development Bank

Otis, J. (1997). Education Crisis *Latin Trade*. Vol.5:(8), pgs. 32-37.

Prensa Libre Newspaper (2001). *Alfabetización entre sí y no*. Electronic Edition, February 22.

Prensa Libre.com

REMHI, (1998).*Guatemala: Nunca Mas. Informe Proyecto Interdiocesano de recuperación de la memoria histórica*. Archbishop's Office for Human Rights. Guatemala City: Author.

Revista Crónica, (1997). *Guatemala en números*. Special Issue, pp. 17-18.

Revista Proceso, (1998). *Cuando la educación se convierte en negocio* [When schooling becomes a business]. 12:4-7.

Rodas Martini, P. (1998). Urge la revolución Pre-primaria. *El Periodico* Newspaper, March 3 p. 11.

Rose, S.D. & Brouwer, S. (1990). The export of fundamental Americanism: U.S. Evangelical education in Guatemala. *Latin American perspectives*. 17, 4:42-56.

Ruano de Flores, E. (1997). *La calidad de la educación*. Guatemala: ASIES Publications.

Ruano, C.R. (2002) Reforma educativa en sistemas administrativos con características premodernas: el caso de Guatemala. *Revista Electrónica de Investigación Educativa, 4* (1). Internet Document:

http://redie.ens.uabc.mx/vol4no1/contenido-ruano.html

Ruano, C.R. (2001) La participación de las Minorias Nacionales dentro de sistemas educativos Pre_Modernos: El caso de los Garifunas de Guatemala. *Eduation Policy Analysis Archives*. 9(23). Retrieved July 12, 2003 from http://epaa.asu.edu/epaa/v9n23.html

Ruano, C.R. (1999). *An Examination of centralized educational policy processes and their impact in Guatemala City, 1976-1995*. Doctoral Dissertation, The University of Toronto. Ottawa: National Library of Canada. http//:www.nlc-bnc.ca

Ruano, C.R. (1997). Living the truth in Guatemala. *Outlook Magazine*. (35) 3: 8,30.

Savedoff, W.D. Ed. (1998). *La organización marca la diferencia. Educación y salud en América Latina*. Washington, DC: InterAmerican Development Bank.

Silva, A. (1996). The politics and techniques of implementing policy innovations: Cross-case analysis and conclusions [pp 139-165]. In *Implementing policy innovations in Latin America*. Antonia Silva, Ed. Washington, DC: InterAmerican Development Bank.

Siglo XXI Newspaper (1999). *Maestros de San Juan Sacatepequez no reciben sus salarios desde febrero*. October, 4th p.13. Author.

Smith, C.A. (1990). The militarization of civil society in Guatemala: Economic reorganization as a continuation of war. *Latin American perspectives*. 17, 4:8-41.

United Nations System (1999). *Guatemala: el rostro rural del desarrollo humano*. Author

United Nations System (1998). *Guatemala: los contrastes del desarrollo humano*. Author.

UNESCO [United Nations Education Science and Culture Organization] (1991). *World education report*. Paris: Author.

Vásquez, B. (1998). Que los investiguen! *Prensa Libre* Newspaper, (January 11th, p. 12).

Vos, R. (1996). Educational indicators; What's to be measured? *INDES Working Papers*. Washington, DC: Inter-American Development Bank.

World Bank: (2000). Poverty and policy in Latina America and the Caribbean. *World Bank Technical Paper No. 467*. Washington, DC: Author.

World Bank (1997). Crime and Violence as Development Issues in Latin America and the Caribbean. The State of Rio de Janeiro and The Inter-American Development Bank. *Seminar on The Challenge of Urban Criminal Violence*. Rio de Janeiro, Brazil, March 2- 4.

World Bank: (1995). Guatemala, basic education strategy: equity and efficiency in education. *Latin America and the Caribbean regional office report No. 13304-GU*. Washington, DC: Author.

World Bank (1978). *Guatemala: Economic and social position and prospects*. Washington, DC: Author.

About the Author

Carlos R. Ruano
El Bosque University
Bogota, Colombia

Email: plaza.ruano@utoronto.ca
carlruano@yahoo.ca

Carlos R. Ruano is Senior Education Specialist, Americas Branch with the Canadian International Development Agency CIDA-ACDI in Hull, Canada. He is also Associate Professor of Educational Foundations at El Bosque University in Bogota, Colombia. He obtained B.A. degrees in Linguistics and History from the University of Ottawa as well as an M.Sc from Georgia State University and Doctor of Education from the University of Toronto. Dr. Ruano's interests deal with the formulation and implementation of educational policy in multicultural and multilingual societies from a comparative and international perspective. The views expressed in this article do not necessarily represent those of CIDA-ACDI.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 662 times since July 17, 2003

Volume 11 Number 22 **July 17, 2003** **ISSN 1068-2341**

Local Impact of State Testing in Southwest Washington

Linda Mabry
Jayne Poole
Linda Redmond
Angelia Schultz
Washington State University Vancouver

Citation: Mabry, L., Poole, J., Redmond, L., Schultz, A. (July 18, 2003). Local impact of state testing in southwest Washington. *Education Policy Analysis Archives, 11*(21). Retrieved [date] from http://epaa.asu.edu/epaa/v11n22/.

Abstract

A decade after implementation of a state testing and accountability mandate, teachers' practices and perspectives regarding their classroom assessments and their state's assessments of student achievement were documented in a study of 31 teachers in southwest Washington state. Against a background of national trends and standards of psychometric quality, the data were analyzed for teachers' beliefs and practices regarding classroom assessment and also regarding state assessment, commonalities and differences among teachers who taught at grade levels tested by the state and those who did not, teachers' views about the impact of state assessment on their students and their classrooms, and their views about whether state testing promoted educational improvement or reform as intended. Data registered (1) teachers' preferences for multiple measures and their objections to single-shot high-stakes testing as insufficiently informative, unlikely to promote valid inferences of student achievement, and often distortive of curriculum and pedagogy; (2) teachers' objections to the state test as inappropriate for nonproficient speakers of English, for students eligible for special services, and for impoverished students; and (3) teachers' preferences for personalized assessments respectful of student circumstances and readiness, rather than standardized assessments. Teachers' practical wisdom thus appeared more congruent than the state testing program with measurement principles regarding (1) multiple methods and (2) validation for specific test usage, including usage with disadvantaged subgroups of test-takers. Findings contrasted a distinction of emphasis: state focus on "*testing* students" as distinct from teachers' focus on "testing *students*."

By 2001-02, standards and standards-based testing were being implemented in 49 states to evaluate school and student performance (Meyer, Orlofsky, Skinner & Spicer, 2002, p. 74), all save Iowa where the state requires district standards (Neuman, 2002) and where it has been reported that virtually all school districts administer the Iowa Test of Basic Skills (ITBS) (Bond, Braskamp, van der Ploeg & Roeber, 1996; Mabry & Daytner, 1997). Formal purposes for standards-based state testing programs typically include statements of intent to improve student learning. For example, Substitute Senate Bill 5953 (SSB 5953), the origin of Washington state's current testing program, opens with these words:

> If young people are to prosper in our democracy and if our nation is to grow economically, it is

> imperative that the overall level of learning achieved by students be significantly increased. To achieve this higher level of learning, the legislature finds that the state of Washington needs to develop a performance-based school system. . . . [T]he state needs to hold schools accountable for their performance based on what their students learn. . . . [I]t will be necessary to set high expectations for all students, to identify what is expected of all students, and to develop a rigorous academic assessment system to determine if these expectations have been achieved. (Washington State Senate, 1992, pp. 1-2)

A decade after implementation of this legislation, has Washington's accountability plan had the intended effect? Have the state content standards, the Essential Academic Learning Requirements (EALRs), and the standards-based test, the Washington Assessment of Student Learning (WASL), improved student learning? With awareness that there have been few empirical studies of the effects of the standards movement nationally (Swanson & Stevenson, 2002) and with particular interest in the local context, an interview study of 31 teachers was undertaken in 2001-02 to discover the impact of reform-oriented, standards-based state testing in southwest Washington, with emphasis on whether it had encouraged changes in classroom practices which promoted improved learning.

Context

National education reform

The implicit theory of action underlying test-driven accountability systems is that testing will improve student learning through provision of accurate data supporting valid interpretations of student achievement, with scores used to identify those who will receive rewards and sanctions, ultimately motivating improved teaching and learning (Baker, 2002). This theory implies that teachers and students are extrinsically motivated, that test scores and the rewards and sanctions they trigger are motivating in the manner intended, and that teachers and students are not working as hard as they could and should (Elmore, 2002). When the nodes in this chain of logic are examined in sequence (see Figure 1), it becomes clear that threats to any link in the chain can result in testing that not only does *not* improve learning but may even be counterproductive.

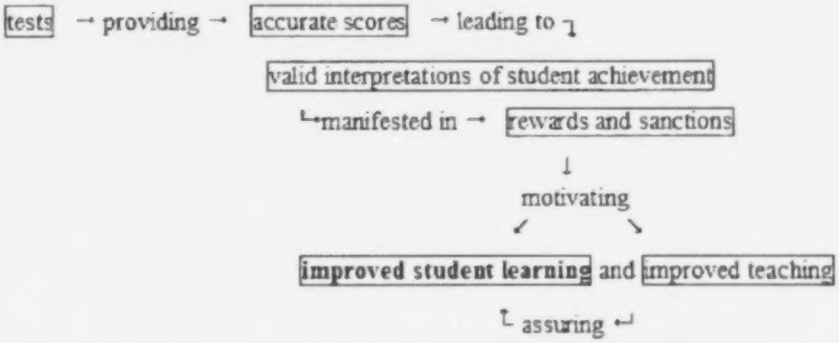


Figure 1. Theory of action in high-stakes state testing and accountability, graphic based on presentations by Baker, 2002 and Elmore, 2002

For example, what if test scores do *not* provide accurate data but if, as has sometimes been charged, the tests are biased against racial and ethnic minorities, females, or the poor? What if rewards and sanctions do *not* motivate teachers to improve teaching but, rather, motivate them to subvert and distort their practice through teaching to the test or "multiple-choice teaching" (Smith, 1991, p. 10)? While the theory of action suggests the mechanisms and sequencing through which testing can improve teaching and learning, it simultaneously suggests the critical junctures at which testing can undermine teaching and learning.

In high-stakes testing, theoretical implications matter much less than real-life implications. Empirical data indicate that scores do tend to rise in the years following the implementation of a new test (Linn, 2000), consistent with the theory of action. Washington state's test data also exhibits this trend (see Table 1), although not uniformly. But whether the higher scores reflect increased student learning is unclear (Haladyna, Nolen & Haas, 1991; Mabry, Aldarondo & Daytner, 1999; Shepard & Smith, 1988; Smith & Rottenberg, 1991). Are the scores accurate, and are they triggering appropriate consequences that yield improved teaching and learning?

Table 1
Trends in Washington state test scores, 1997-2001: Percentages of students meeting state standards in reading, math, and writing, based on data available online at www.k12.wa.us

	Scores by years

Test subjects and grades	1996-97	1997-98	1998-99	1999-2000	2000-01
Reading					
grade 4	47.9	55.6	59.1	65.8	66.1
grade 7		38.4	40.8	41.5	39.8
grade 10			51.4	59.8	62.4
Mathematics					
grade 4	21.4	31.2	37.3	41.8	43.4
grade 7		20.1	24.2	28.2	27.4
grade 10			33.0	35.0	38.9
Writing					
grade 4	42.8	36.7	32.6	39.4	43.3
grade 7		31.3	37.1	42.6	48.5
grade 10			41.1	31.7	46.9

Bar graph based on *fourth grade reading scores* (first row in the table above), rounded to the nearest whole number, to visualize score increases more clearly

Scores					
68					
67				65.8	66.1
66					
65					
64					
63					
62					
61					
60			59.1		
59					
58					
57		55.6			
56					
55					
54					
53					
52					
51					
50					
49	47.9				
48					
47					
46					
Years	1996-97	1997-98	1998-99	1999-2000	2000-01

The consequences of state testing in the U.S., where the stakes are high and getting higher, indicate widespread acceptance–at least implicitly–of the theory of action. Currently, 43 states require school report cards (including Washington), with two more in development, and 20 of these require that the report cards be sent home to parents. Twenty states (not including Washington) have the authority to impose serious sanctions on low-performing schools: school closure or reconstitution, student transfers, and loss of funding; three more states will be able to do so within two years. Eighteen states (not including Washington) provide rewards to high-performing or improved schools, with two more set to do so within two years. Fifteen states use test scores alone, with no additional evidence, to evaluate schools (Meyer, Orlofsky, Skinner & Spicer, 2002).

The difficulty charter schools are experiencing in trying to raise test scores (e.g., Gewertz, 2002) is heightening awareness that raising test scores in straightened educational circumstances is not easy. Perhaps because of this, test-triggered stakes are increasingly being borne by students who are relatively defenseless (Elmore, 2002). In particular, in seventeen states, a number that will increase by seven in the next two years, adolescents cannot graduate from high school without passing exit or end-of-course exams (Washington will require a graduation test in 2008). An elementary or middle school child's promotion to the next grade is contingent on test scores in four states (not including Washington), a number that will double in the next two years. Remediation is required for students failing promotion, end-of- course, or high school graduation exams in seventeen states, most but not all of which provide funds for the remedial instruction (Meyer, Orlofsky, Skinner & Spicer, 2002).

The newly reauthorized Elementary and Secondary Education Act (2001), dubbed "no child left behind" (NCLB) and sometimes derisively called "no child left untested," furthers the trend toward more state testing and higher stakes. Stakes include federal Title 1 funding and now, for underachieving schools, requirements to provide school choice to parents in year 2 of a school's continuing low test scores, tutoring with parental choice as to providers in year 3, replacement of curriculum and/or staff in year 4, and reconstitution in year 5. The basis of these sanctions is state test scores. In Washington, few schools currently meet NCLB standards: only 36 of 1162 elementary schools, 19 of 554 middle schools, and 13 of 505 high schools (*Oregonian*, 2002).

Superseding the Goals 2000 call for a national system of tests in 1994, (Note 1) NCLB requires increased state testing, including standards-based assessments of reading and math for all students in grades 3-8. In order to receive Title 1 funds, the law requires attainment of proficiency by all students–including minorities, students with limited proficiency in English, and low SES students–within twelve years and proportional annual yearly progress (AYP) in the interim. To discourage states from using easy tests that might distort achievement or lower expectations, the law also requires that scores on state tests be confirmed against scores on the National Assessment of Educational Progress (NAEP). (Note 2)

The AYP targets are about double the score increases empirically documented by NAEP over time, which suggest that it might take not twelve but more than 100 years, by optimistic estimate, to reach the required 100% proficiency. The AYP targets have been judged especially "unrealistic" for schools and districts where small enrollments of disadvantaged subgroups of students will result in statistically unstable results (Haertel, 2002). The targets also appear painfully unrealistic for chronically under-resourced urban schools (Lewis, 2002; Yakimowski, 2002). National policy thus exhibits confidence in a theory of action that is empirically suspect.

State policy

The standards-based, test-driven educational reform initiative mandated by the legislature in SSB 5953 in 1992 lists four purposes for the state of Washington's accountability system:

- to assess students' academic learning
- to evaluate instructional practices
- to select students for remediation
- to hold schools accountable for student learning (Washington State Senate, 1992, p. 10).

These are very similar to the four purposes for assessments recently listed by Shepard (2002)–diagnosis, monitoring, student selection, and program evaluation–with the warning that making a test more valid for one purpose might make it less valid for a competing purpose. Frequent similar admonitions from the measurement and evaluation communities indicate that multiple purposes for a single test are usually problematic, as different purposes are often in unwitting conflict, undermining achievement of any of the goals (e.g., Mabry, 1999). For example, tests used for school accountability have often proved vulnerable to "score pollution" as school personnel administering the tests succumb to pressure to raise scores through a variety of means, some ethically, legally, or statistically questionable (Haladyna, Nolen & Haas, 1991; Haney, 2000; Linn, 2000; Sternberg, 2002).

Score increases are not always credible, as evidenced by discrepancies between some state NAEP scores and scores on the state test (e.g., Haney, 2000) and by the so-called Lake Wobegon effect–states' insistence that more than half of their students were "above average" (Cannell, 1987), a statistical impossibility. As educators scramble to raise scores to protect their schools, students, and themselves from high-stakes penalties, improved state test scores may not necessarily reflect improved student achievement. Inflated scores would obstruct understanding of students' academic learning and would obstruct identification of students needing remedial assistance–two goals of Washington state's accountability system.

The Washington Assessment of Student Learning (WASL) tests literacy and math at grades 4, 7, and 10 and offers multiple-choice and constructed-response items, both short and extended writings. Described as a criterion-referenced assessment aligned to state standards (Meyer, Orlofsky, Skinner & Spicer, 2002, p. 75), the WASL is administered in late Spring. Student performance is judged to be "above standard," "meets standard" (the required level of proficiency), "below standard," and "well below standard." In 2001, schools and districts were required to reduce by 25% the number of students not meeting the state's required standard and to include in public reporting their goals and plans to do so (online at the state education agency's website, www.k12.wa.us). As noted, in comparison to some states, the stakes associated with the WASL are relatively low: schools are not threatened with closure or reconstitution; funds are not withheld because of low scores; students in grades 4 and 7 are not retained at grade level or compelled into remedial education if they do not meet standards; high school students' eligibility for graduation will not be contingent upon WASL scores until

2008 (Note 3) (Meyer, Orlofsky, Skinner & Spicer, 2002, pp. 74-75).

Is Washington's testing program having the intended effect, assuring that "the overall level of learning achieved by students be significantly increased"? State statistics generally suggestimproved achievement (see Table 1) but, as of 2001, national statistics indicated that less than a third of Washington's fourth- or seventh- graders had scored at the "proficient" level on NAEP reading, writing, math, or science tests (Orlofsky & Olson, 2001). Of course, it might be that the scores reflect state learning goals but not national learning goals. It might also be that the state's standards-based testing program is improving learning but not yet measurably since, elsewhere, indications have been found that state reforms are resulting in teachers' adoption of classroom practices consistent with standards (Swanson & Stevenson, 2002). Local evidence of teachers' acceptance of Washington's state standards and of positively evolving classroom practices, if occurring, might suggest gradual improvement which could become measurable in the future.

The research reported here investigated the resonance between state testing and classroom assessments, whether feedback from the WASL helps teachers understand their students' achievements and plan more effective learning opportunities, whether local classroom practices are changing, whether state testing is encouraging the alignment of curriculum to state standards and, if so, whether the alignment is educationally beneficial.

Method

The approach to the study undertaken in Fall 2001 was qualitative, subscribing to a view of human phenomena as socially constructed (Vygotsky, 1978) from individuals' perceptions of reality. The research process adhered to interpretive research traditions and methods respectful of emergent design, multiple perspectives, and inductive analysis (Denzin, 1989, 1997; Denzin & Lincoln, 1994; Erickson, 1986; Mabry, 2002; Merriam, 1998; Stake, 1978; Wolcott, 1994). Two data collection methods were employed: review of documents (Hodder, 1994) related to testing in Washington state and, more importantly, semi- structured interviews (Fontana & Frey, 1994; Rubin & Rubin, 1995) of practicing teachers in the local area.

After approval of the study by a university Institutional Review Board and signed consent from each interviewee, graduate students at Washington State University Vancouver (Note 4) interviewed 31 local teachers in Fall 2001. The sampling strategy was purposeful rather than representative or randomized, with each graduate student identifying and interviewing two teachers who taught a subject at a grade level of specific interest to the interviewer. (Note 5) This subject selection strategy maximized the sensitivity of the interviewers to each teacher's subject area and grade level.

Of the 31 teachers interviewed, 19 taught in high schools, 5 in middle schools, and 7 in elementary schools (see Table 2). Their teaching experience totaled 547 years, with an average of 18 years each and a range of 1-40 years. Nineteen interviewees were female and 12 were male. All of the teachers' schools were located in southwest Washington state, and all but one of these was a public school. The teachers included 13 who taught subject areas and grade levels tested by the WASL and 18 who did not. Of the teachers whose students were tested on the WASL, 9 taught in high schools, 2 in middle schools, and 2 in elementary schools.

Table 2
Teachers interviewed, the subjects and grade levels they taught, and whether these subjects and grades levels were tested using the WASL (n = 31)

Level	Subject/grade	Tested subject at this grade level?	Teachers (by pseudonym) of this subject at this grade level and years of experience
High school *n=19,* *9 in tested grades*	English-language arts*	YES	Ms. Apple, 3 years Ms. Brush, 15 years Mr. Carr, 7 years Mr. Dustin, 20 years Ms. Hand, 22 years Ms. Kroner, 7 years Mr. Twain, 25 years Ms. Underwood, 20 years

	mathematics**	YES	Mr. Alder, 19 years
	science	no	Mr. Liu (biology), 17 years Mr. Ming (biology), 17 years Mr. Ochre (biology), 9 years Ms. Vargas, 20 years Ms. Walker, 20 years Mr. Banks, 34 years
	family and consumer ed	no	Ms. Crane, 30 years Ms. Doe, 14 years
	foreign language	no	Ms. Good, 22 years
	social studies	no	Mr. Inder, 1 year
Middle school	English-language arts	YES	Ms. Frank, 12 years Ms. Nunn, 5 years
n=5, *2 in tested grades*	history	no	Mr. Eggle, 25 years
	social studies and other	no	Ms. Grant, 18 years
	grade 6	no (private school)	Ms. Smith, 16 years
Elementary school	grade 1	no	Ms. Park, 16 years
	grade 2	no	Ms. Hallo, 13 years
n=7, *2 in tested grades*	grade 3	no	Ms. Jones, 30 years Ms. Quinn, 40 years
	grade 4	YES	Ms. Roberts, 21 years Mr. Exeter, 8 years
	grade 5	no	Mr. Felix, 21 years

* One teacher taught English-language arts and also civics and philosophy.

** This teacher taught math and also P.E.

A collaboratively constructed interview protocol (see Exhibit 1) guided semi-structured interviewing (Fontana & Frey, 1994; Rubin & Rubin, 1995). Interviews lasted approximately 45 minutes each. Interviewers attempted to capture as many direct quotations as possible, with some interviews tape-recorded with permission of the interviewees and others recorded in hand-written notes typed up soon thereafter. For purposes of developing a high-quality database with strong internal validity (Campbell & Stanley, 1963) or descriptive validity (Maxwell, 1992), a comprehensive validation strategy (Mabry, 1998) was used, with each interview written up and presented to the interviewee with a request for review, correction, and elaboration.

Exhibit 1. Protocol for semi-structured interviewing of teachers

- How many years have you been teaching? What grade levels and subject areas have you taught? Has all of your teaching occurred in the state of Washington?
- How do you assess your students' achievement?
- How did you develop your approach to student assessment? Why did you take this approach? What influenced your thinking? How long did it take to develop? How has it evolved over time (if it has)?
- Have you had training in assessment? If so, how much training have you had? How would you describe the type of training you have had? Has your assessment training been related to specific content areas?
- As the state has developed requirements for student learning and for assessing student achievement, has your teaching changed? If so, what has changed about your teaching? Do you consider the changes to be improvements?
- How do you feel about the WASL (Washington Assessment of Student Learning)? Why?
- How do you prepare your students for your assessments (if you do)? How do you prepare them for state assessments (if you do)?
- If you were to change your classroom assessments, what would you like to do differently? If state assessments were to change, what type of change would you favor?
- Does your school or district require testing (other than state testing)? Are tests part of your school's or district's graduation requirements for high school students?
- Is there anything you would like to add?

Thank you very much for your time and information! I will type up my notes from this interview and give them to you. I would very much appreciate it if you would read the notes and make any corrections to improve accuracy. If there is anything you would to add at that time, I hope you will feel free to make additions then. Again, *many thanks*!

Data analysis was emergent in character, with meaning sought in the data without reference to *a priori* categories (Denzin & Lincoln, 1994; Erickson, 1986; Mabry, 2002; Wolcott, 1994). Analysis involved four phases and two validation efforts. In the first phase, pairs of graduate students analyzed their four interviews for patterns, including commonalities and distinctiveness across their four subjects. This thematic content analysis (LeCompte & Preissle, 1993; Miles & Huberman, 1994) and the resulting preliminary interpretations were written up in eight separate preliminary reports. In the next phase, the first author conducted a similar content analysis across the eight student reports, identifying 29 themes overall and grouping them in four emergent categories: (1) classroom impact, (2) student impact, (3) teacher impact, and (4) teachers' perspectives (see Table 3).

Table 3

Themes emerging from content analysis of teacher interview data, identified from eight preliminary interview reports and grouped into four categories

		Interview reports							
	Themes	1	2	3	4	5	6	7	8
A	**Classroom impact**								
	Teachers' approaches to classroom assessment	X	X	X	X	X	X	X	X
	Training in assessment for teachers	X	X	X	X	X	X	X	X
	Changes in assessments over time		X			X		X	
	Usefulness of the WASL for classroom practice					X			
	Impact of state standards/tests on curriculum and instruction (or resistance to impact)	X	X	X	X	X	X	X	X
	Preparation in class for the state test		X	X	X	X	X	X	
	Impact of the WASL on classroom assessments			X			X		
	Impact of the WASL on classroom environment					X			

B	**Student impact**								
	Student accountability based on WASL scores (e.g., graduation, retention)	X		X			X		X
	Equity to students, including						X		
	• Students whose first language is not English			X	X				
	• Special education students			X	X				
	• Transfer students				X				
	• Minority students			X	X				X
	• Students in difficult circumstances (including SES)			X					
	Impact of the WASL on students' self- esteem/stress/anxiety		X	X		X		X	
C	**Teacher impact**								
	School/teacher accountability based on WASL		X	X	X	X	X	X	X
	Pressure to perform well on the WASL			X	X	X			
	Impact of the WASL on teacher professionalism			X					
	Contrasts of interest to us, including								
	• Public and private schools					X			
	• Tested and non-tested grades/subject areas by the WASL	X		X	X	X	X	X	
	• Teacher assessments and state assessments	X	X	X	X	X	X	X	X
D	**Teachers' perspectives**								
	Teacher approval or What teachers like about state testing		X			X	X		X
	Teacher disapproval or What teachers do not like (or would change) about state testing		X	X		X	X		X
	Questioning the constructs tested (or that should be tested) by the WASL	X	X	X		X	X	X	
	Questioning the difficulty level of the WASL				X	X			
	Scoring concerns		X				X	X	
	Questioning the expense of testing							X	

(Note. An **X** indicates that data related to the theme (listed by row) were found in the preliminary report (listed by column) in phase 2 of data analysis. Further review in phase 3 identified additional sources of data on these themes, revisions to these themes, and additional themes.)

A third data analysis phase involved micro-review by the authors of the entire data set for comprehensive identification of all data points related to each theme and category. The final phase of analysis was the identification, drafting, and formalization of findings. A draft of the resulting manuscript was offered for review and critique to all 16 interviewers in a second validation effort.

The data and findings were structured for reporting according to the four major thematic categories. The teachers quoted are identified by pseudonyms.

Classroom impact

Consistent with other findings about the impact of standards-based reform (Swanson & Stevenson, 2002), the data made clear that all of the teachers interviewed were highly aware of the state reform initiative and that state policy was definitely felt in local schools and classrooms. Veteran teachers stood as witnesses to changes ushered in by the reform efforts. For example, a teacher with forty years' experience observed, "Early in my career, there was very little emphasis on assessment. This has changed, most recently because of the state Essential Learnings" (Ms. Quinn). She, among others, indicated that she had seen public school assessments evolve over the years from reliance on intuitive teacher judgments to formal state standards, the Essential Academic Learning Requirements or EALRs.

Response to state initiative

In response to Washington's state standards and the state test, the WASL, most interviewed teachers had adjusted their classroom practice, they reported, some more than others. While most teachers said their instructional styles had not changed, many said the content they taught had altered, consistent with other research indicating that teachers feel state frameworks are redefining curricula (Shore, 2002). Some teachers reported positively that new state standards provided explicit objectives which helped focus their teaching. For example, one said, "It gives students and teachers a target" (Ms. Frank). Another commented, "If you are going to try to do the job right, you should always have [the EALRs] beside you so you can see what you're doing is on target" (Ms. Good). These teachers had approvingly accepted the EALRs as their teaching goals. A high school science teacher suggested that acceptance might sometimes have been compelled rather than willing:

> In general, [the WASL] is a good idea because it has forced people to be accountable. Many kids from middle school didn't have the basics, and we had to spend time re-teaching what they should already have known. (Ms. Vargas)

Some teachers expressed ambivalence or frustration regarding the superimposing of state goals over their own aims and approaches. For example, a high school teacher expressed hope that aligning her teaching to the EALRs had "tightened up" her teaching but also said she found it frustrating when this forced elimination of her successful hand-crafted units: "It's hard to ditch your pet projects" (Ms. Good). A third grade teacher who had previously taught thematically said she now taught subject-by-subject, with special attention to state content standards, a change she described not as an improvement but as "a necessity in the constantly changing world of education" (Ms. White).

With the state graduation test postponed until 2008, classroom practices were less affected at the high school level, according to the teachers, and less evident in content areas which were not yet tested by the WASL. Even so, many high school teachers indicated strong impact of the test on their practices. For example, the only interviewee who baldly admitted to teaching to the state test was a high school teacher who said he had been directed to do so by his principal because there was “a lot at stake” (Mr. Ochre). Greater impact was apparent at the elementary level, particularly in fourth grade, a tested grade. At this grade level, wholesale displacement of the curriculum was noted by some, including one fourth grade teacher who said, “Teachers in [my] building spend from about November to mid-April focused on the WASL” (Ms. Roberts).

Classroom assessments

Every teacher reported using a variety of assessment methods and techniques in the classroom. Often, these featured performance assessment, their reported practices ranging from observations of student performances to portfolios and projects. Variations among the teachers' assessment ideologies and practices suggested adaptations harmonious with personal style, the range surprising some interviewers. For example, Ms. Hand and Ms. Kroner were described by their interviewer as having "vastly different takes on what constituted appropriate assessment, yet both were excellent teachers who were obviously very dedicated to their profession." Even teachers who taught the same subjects and grade levels approached assessment differently (e.g., Mr. Twain and Ms. Underwood, high school English; Ms. Vargas and Ms. Walker, high school science). Suggesting adaptations based on experience and changes in student populations, teachers consistently described continuous efforts across time as “constantly evolving each year as my class changes and the world around them changes” (Mr. Banks).

Using assessment to ensure student success, rather than to identify weaknesses for remediation or penalty, emerged as an important distinction between classroom and state assessments in comments from some teachers, such as:

> I want kids to be successful in my classroom. I'm not there to fail students. I'm there to teach students. [For] those with low academic abilities, if you put too much emphasis on testing, you would see a high failure rate. (Mr. Liu)

Another who offered an earnest rationale for using assessment to improve achievement rather than to punish students said:

> Over my 17 years of teaching, I've really changed my approach to student assessment. Initially, I started out really being worried about content. The value of their grade was based more on testing–maybe 80%, 90%–less on what they did in the classroom, less on behavior. Over the 17 years, I've changed that. Maybe it's not so important *what* they're learning but *how* they're going about doing it, how they're approaching what they're doing in class. I've shifted my emphasis from content and techniques to behavior and work-related skills. . . .
>
> I want kids to be successful in the classroom. If I based [grades] strictly on content, I'd see too many kids failing. I think today more kids are coming to my classroom without the tools needed to be successful in terms of learning the same level of content that I expected 17 years ago. So, should they get slapped again because they're not prepared or not able to do what I expected 17 years ago? I don't think so. . . .

> I think they come with a lot more baggage today than they did 17 years ago–a lot more personal issues, parental guidance issues. . . . We're doing more parenting, and that's just as valuable. (Mr. Ming)

Mr. Liu and Mr. Ming described their assessments as compassionately tailored to the realities of their students' troubled circumstances and consequent skill levels, adaptations not possible with the state's standardized test. Manty teachers spoke of efforts to personalize assessment, and many indicated they wanted to implement even more personalized assessment but were prevented by serious limitations on a critical resource–time.

Overall, classroom impact data tended to agree with prior research indicating that testing is having so profound an impact in many classrooms that reform is driving curriculum, a positive effect to the extent that there may now be "less fluff" but negative where pressure to raise test scores eliminates flexibility (Horn, 2002) and focuses on scores rather than on students.

Assessment training for teachers. The teachers' preparedness to meet formal expectations regarding assessment (Washington State Senate, 1992; AFT, NCME & NEA, 1990) appeared to be uneven and inadequate, consistent with wider reports of insufficient teacher training in assessment (Hargreaves, Earl & Scmidt, 2002; Stiggins & Conklin, 1992). Most interviewees, although not all, considered their undergraduate assessment training inappropriate for classroom use. For example, two described their pre-service assessment training as "minimal" (Ms. Vargas, Ms. Walker). Another said she had had only one assessment class in college, which proved unrelated to her content area and which she found to be "useless" for her own teaching (Ms. Grant).

While two teachers reported no assessment training whatever since their initial teacher preparation and many said they had not taken post-graduate assessment courses, others described in-service training in assessment as a "never-ending process" (Ms. Jones) of classes, meetings, seminars, and workshops for local educators and administrators. However, opinions about the quality of professional development in assessment were sometimes no more positive than those regarding college and university assessment courses, one teacher describing assessment training as "a huge inadequacy" (Ms. Nunn). Not only the adequacy but also the appropriateness of the training offered to teachers emerged as suspect. Recent assessment training by one local district, some teachers said, emphasized writing WASL-like questions for implementation in their classrooms "to get [students] used to that type of assessment" (Ms. Park), rather than understanding of measurement principles. In-service training had preempted rather than promoted teacher-developed assessments, reported one teacher who said, "I never developed my own assessment techniques because I was trained by the school and district in the way that they wanted assessment done" (Ms. Apple).

Some teachers identified their colleagues as a more important source of assessment information that pre-service or in-service training. Two teachers referred to assistance they had received from mentor teachers (Mr. Eggle, Ms. Frank), and one of these derided "new assessment ideas [as] just old ideas draped in new jargon that confuse and threaten older teachers" (Mr. Eggle).

Student impact

The teachers' awareness of the impact of the state test on their students was abundantly evident in their comments. Some indicated that they considered student accountability a commendable state goal. One approving teacher, for example, said, "The WASL is a good thing to hold kids accountable" (Ms. Park). A teacher with a less favorable view of the WASL nevertheless implied that more state testing for the purpose of making promotion and retention decisions was desirable, with a lament that "we do not have any exams to hold kids accountable for moving on to the next level" (Mr. Ming). Another teacher indicated preference for earlier and more frequent imposition of test-based consequences for students in commenting on the inappropriateness of delaying student accountability until high school graduation, saying that he considered it "odd" that the WASL "counts" only for tenth- graders (Mr. Carr). Most teachers, however, expressed serious concerns regarding the WASL's impact on students, as the comments to follow indicate.

Effects of the state test on student self-esteem

Of the teachers interviewed, those who taught at a tested grade or in a tested subject and those who did not both expressed concern regarding the impact of the WASL on student self-esteem. The teachers typically described environments in schools and in classrooms as highly charged during testing windows, one teacher referring to "a lot of stress for both the takers and the administerers of the test" (Ms. Good). Even a teacher who spoke favorably about the test warned that the WASL:

> does have too much pressure and overwhelms the students. The scores affect their self-esteem. Nothing is in place [for students who don't meet] the standards. (Ms. Park)

Developmental appropriateness. Data indicated that few teachers considered the state test too easy relative to the content and expectations of their classrooms. One who did said, "[W]hat we expect from students is a lot harder than anything that the WASL tests for" (Ms. Apple). Much more common were concerns that the test was too difficult for some students, to the point of being "not developmentally appropriate for fourth graders. I've seen kids crying about the nightmares they've had over this" (Mr. Felix). A fourth grade teacher expressed the greatest degree of concern about the test, saying:

> It is developmentally too difficult for [fourth grade] students. They try so hard when they have no chance of passing. [In the writing portion of the test,] to get a four [the highest score, requires a student to write] better than I could write. This is what they call raising standards. Raising standards means putting it beyond their developmental level and hoping they are going to reach for it. We know that doesn't work. (Ms. Roberts)

The difficulty level of the fourth-grade test and perceptions of its developmental inappropriateness led some teachers to the conclusion that the test was unfair. Perceptions of inequity were exacerbated for students eligible for special services, for English language learners, and for students with low socioeconomic status.

Effects of the state test on diverse and disadvantaged students

Testing special education students. Some teachers suggested that, for special education students, test time was ill-spent because the WASL offered them "no chance" to demonstrate their knowledge and skills. Even a teacher who approved the test said, "I feel that special ed students should not have to take the test . . . Their time could be better spent on more educational experiences" (Ms. Park). Another teacher objected:

> [M]y EMR, which is educable mentally retarded, students have to take the WASL. Learning disabled students have to take the WASL. . . . The EMR students are not going to be successful [on the test], yet we put them through 400 minutes of sweat when they could be having other kinds of experiences. (Ms. Roberts)

Few students could be exempted outright from taking the test, teachers said. However, accommodations were available for students classified as eligible for instructional assistance, but only if the accommodations provided during testing matched ordinary classroom accommodations. Although this general policy sounded reasonable in the abstract, specific restrictions on accommodations rendered it useless, according to one teacher who said:

> If a person is learning disabled in writing, and we wanted someone to scribe for them–take dictation–[the student] would have to have that person all year long, every time we had a writing assignment. . . . It can't be just an accommodation for the WASL testing window. Individual students [would] be requiring a lot of time from our [teaching assistants], which we don't have. (Ms. Roberts)

Other accommodations teachers thought might permit documentation of actual achievement were sometimes denied, as described by one teacher recalling a "special needs student [who] could sit down at the computer . . . [where] she had a way of expressing herself" but was not allowed to use the computer when taking the WASL (Ms. Crane).

These teachers' experiences of testing special education students reflected the views of teachers nationally who have objected to state tests as merely providing a new way to show these students they are failures (Horn, 2002). Elsewhere, such perspectives have been brought to bear in legal action, for example, in the 1998 class action lawsuit charging the Indiana state test with unfairness to special education students.

Testing English language learners. The teachers described testing practices for English language learners as no better than those for special education students. A middle school teacher observed:

> ESL students only get a one-year exemption from the WASL, which is not nearly enough [time] to [become] familiar with the language, material, and culture to do well on the test. (Ms Nunn)

A fourth-grade teacher fretted:

> They test everyone including kids who have only been [in the U.S.] for a year and a half, so they're taking a test they cannot read. . . . Even though you [might] say, "Oh, they can have assistance," the ESL kids [can only] have the problems read to them verbatim. (Ms. Roberts)

The students would still have to write answers in English.

One teacher who declared that the WASL "doesn't work well when used to assess minorities or special ed students" raised a question of serious practical consequence: "What do we do with the students who cannot pass [the test] year after year and fail to advance?" (Mr. Ochre).

Testing low SES students. Teachers indicated that they considered students in straightened economic and personal circumstances in no less need of consideration than special education students and English language learners. One teacher pointedly predicted, "I'm sure the WASL scores will be best correlated to how much does your mom and dad make economically" (Mr. Ming). In fact, as this teacher intuited, historically, the strongest correlate with scores has been socioeconomic status. The effect of socioeconomic status on test scores was no small matter to the teachers interviewed. For the three-year period 1998-2000, 9-10% of the population in the state of Washington had been considered impoverished (U. S. Census Bureau, 2000). At the time of this study, Vancouver, in the Portland metropolitan area, was suffering from Oregon's highest unemployment rate in the U.S. (Preusch, 2001).

Some teachers poignantly acknowledged increasing levels of economic and social disarray in many families and the consequent calamity in the lives of stricken students. One teacher worried about "how much assistance and guidance do parents provide and are they abusive or intoxicated" (Ms. Grant). No accommodations were available for students suffering the effects of these and other detriments to their real academic opportunities, and no consideration of such background variables were taken into account in calculating their individual achievements as scores on the state test.

Testing students with diverse learning styles. Consistent with the popular theory of multiple intelligences (Gardner, 1983), some teachers noted an inequity derived from discrepancies between the test (content and format) when compared to the different kinds of skills, achievements, and knowledges students might actually possess. Like most standardized achievement testing, the WASL emphasized "logical- mathematical" knowledge and skills over most other types of achievement. Within this theoretical context, these teachers implied that students with strong accomplishments in areas not included on the WASL were unfairly judged non-proficient by a state test that measured a restricted range of achievement.

Teacher impact

Accountability pressures

Most interviewees explicitly recognized that "society wants accountability" (Ms. Good) and that "raising the standard would raise the credibility of the American public school system" (Mr. Carr). They were keenly aware of public scrutiny of WASL scores published in local newspapers.

Almost unanimously, more pervasively than has been reported nationally (Abrams, 2002), the teachers, even those who described themselves as relatively unaffected by testing and test pressures, noted societal pressures related to scores and accountability. Fewer than one-fourth of the teachers interviewed, most of these in untested grades, indicated that the WASL had little impact on them. One said that his plan for avoiding test-related demands was to retire so that he would be "long gone" before it was necessary for him to align his curriculum with the state test (Mr. Liu).

No teacher in this study objected to accountability *per se*, but several teachers expressed frustration at being held accountable for test results when student performance depended not only on teaching but also on factors beyond teacher control, factors they listed as including class size, student ability, primary language, eligibility for special services, socioeconomic status, transience, family difficulties, and motivation. "Your teaching [will] eventually be judged by the kids who blow it off," fumed one teacher (Ms. Good). Another recognized teacher vulnerability where "students perform badly on the WASL intentionally to make a point" of their own objections to the test (Mr. Dustin).

While several teachers considered the impact of state testing meritorious, most expressed concern regarding the appropriateness of the state's prioritization of test scores in reckoning school accountability. This was consistent with other research findings that teachers do not oppose standards or accountability, but most disagree with current uses of test scores for school accreditation (Abrams, 2002; Shore, 2002).

Effects on classroom instruction and assessment

Some teachers in this study described classroom effects similar to reported trends indicating that state tests "deform curricula" (Schoenfeld, 2002). The teachers identified such things as how to fill in the bubbles on answer sheets and how to follow prompts as examples of local WASL preparation activities which took time away from regular teaching and learning.

The extent of curriculum displacement alarmed some teachers, one of whom said, "When we do the WASL, our school is in chaos for the entire time. I lose a month of teaching. It affects the whole school" (Ms. Doe). Others reported that "test prep" consumed as much as five or six months in a tested grade. One teacher complained that the test overwhelmed classroom instruction even in untested grades:

> I guess I'm one of many teachers who feel there is so much emphasis on the WASL that it has almost become the focus of our teaching. I'm not real comfortable with that. . . . [F]or the fourth graders, the minute they enter fourth grade, they're hearing about the WASL and how they have

> to do well on the WASL. . . . But even at third grade, I find myself saying to students, "This is the type of question that you will have on the WASL when you are in fourth grade." . . . [W]e're just so test-oriented that we've kind of lost sight of what education truly is. (Ms. Quinn)

Some teachers expressed resistance to reallocating instructional focus and time for test preparation, one saying, "I can't just prepare my students to take the WASL. It's not the only thing that should be assessed" (Ms. Brush). Another, who complained that the WASL, a standardized test, "doesn't measure anything that we teach our kids," declared, "we are not willing to change because what we do for our kids is what they need" (Ms. Apple).

But many fell into line, some with misgivings or under duress. A high school science teacher, for example, had reluctantly added earth science to her curriculum because it was found in the EALRs, although it was outside her specialty (Ms. Walker). Another teacher reported a shift away from thematic instruction and toward a fragmented approach to the curriculum as she "hit subject matter individually while constantly checking and re-checking the EALRs" (Ms. Jones). One reported her teaching was "becoming more canned" (Ms. Hallo). Another described herself as physically displaced in her own classroom, to some extent, by tutors brought in to ensure her students were prepared for the WASL (Ms. Doe).

The amount of class time devoted to external assessment was not limited to preparation for and administration of the WASL. Teachers reported at least eight additional standardized achievement tests, seven developed and marketed by big-name commercial testing corporations, in use in their schools or districts.

Not only instructional practices, but classroom assessment practices, too, were increasingly pressed into the WASL mold, data indicated. For example, one teacher said she had reorganized her students' portfolios "to match the EALRs." As a member of her district's "assessment training team," led by an official from the state education agency, she was "learning how to write a practice test similar in format to the WASL" as her district developed "a WASL-like practice tests for second graders" (Ms. Park).

Teacher perspectives

Slightly less than half of the teachers interviewed expressed approval of the WASL or of some aspects of it. Two praised the test's emphasis on "process," one adding that this emphasis was "good because we're trying to make a more fair assessment" (Ms. Hand), and the other praising partial credit given to students who showed workable math procedures even when answers were ultimately incorrect (Ms. Doe). The latter also approved the WASL's authentic eliciting of "the same skills [students] use in real life" (Ms. Doe).

Most expressions of approval included qualifications. For example, one teacher said the WASL was "probably a good thing" (Mr. Inder), another that she felt positive "for the most part" (Ms. Frank), and another that it was a good thing to hold students accountable although "improvements to the test" were needed (Mr. Alder). One said the WASL "can assess some [students]. I think that it cannot assess all" (Ms. Roberts). Content limitations were noted by a teacher who observed that the test content was "not the only thing that should be assessed" (Ms. Brush). One teacher who approved the test distinguished between its quality and its utility: "I like the test. I just don't know how it should be used" (Mr. Carr).

The most positive opinions were offered by two teachers involved in developing either a practice WASL-like test or a rubric to standardize the assessment of student writing. One of the two had been a member of a district assessment team for four years, "so long it has become a part of me, and I have begun to buy into it." Even so, her praise of the WASL was qualified: "The test is still new. The kinks have to be worked out" (Ms. Park). The other, who said she had helped develop Six Trait Writing Assessment, indicated that she had changed her teaching in response to the state standards, which she considered congruent with her beliefs and practice, but that she disapproved of the WASL as "much too narrow a device" (Ms. Underwood). This small (n=2) positive correlation between individuals' involvement with test development and their approval of the WASL was consistent with findings from a nation- wide study:

> [T]he promotion of greater receptivity towards change at a local level, which might entail teacher knowledge about the reform, shaping attitudes toward reform objectives, or providing greater "how- to" knowledge instrumental for implementing change. . . . appears to be a likely mechanism through which this policy reform operates. (Swanson & Stevenson, 2002, p. 15)

However, it was unclear whether local data suggested that teachers' close scrutiny of the testing system led them to appreciate the WASL or, alternatively, suggested that involving teachers in development of standards and tests habituated and coopted them.

Most teachers took issue with the test, the most virulent wording coming from one who called the WASL "stupid" (Mr. Banks) and another saying, "I despise it" because of its counterproductiveness regarding learning and its deprofessionalization of teachers (Mr. Twain). Complaints centered on negative impacts to curriculum, students, classrooms, and schools, as previously detailed, articulating questions and concerns about equity and developmental appropriateness, as noted earlier, and about validity, scoring, expense, and volatile state policies and requirements, to be discussed in the next section.

Teachers' objections also included lack of useful feedback in the reporting of test results. "It would be nice for kids to get the tests back and see the mistakes that were made so that they could focus on their weaknesses," said one of two teachers (Ms. Vargas) who objected to delayed notification of WASL results. The other estimated the delay as "six months" after test administration, too late for corrective instruction. When results did arrive, she complained, there were further obstructions to fulfillment of the state goal that testing help identify remediation needs:

> As far as I can tell, there has been no interpretation of what failing test results mean. . . . [And] I am not allowed to keep the test results. I am only allowed to see them for a short time because they are locked up. I don't know if that's in every school or just in this school. (Ms. Roberts)

One teacher objected to the content of a specific item, saying he had "lost respect" for the WASL after publicity about a tasteless question that referred obliquely to a notorious trial involving a teacher's alleged seduction of a student (Mr. Ochre).

Overall, local teachers' concerns about state testing closely matched those of their colleagues nationally: fairness, timeliness of feedback, diagnostic value of test result reports, single-shot testing, pacing in classrooms, the number of tests, extraneous factors that affect scores, and pressure to cover all the standards (Shore, 2002).

Validity concerns

Validity through multiple measures. Although no teacher used technical terms in responding to interview questions, analysis of data from the perspective of traditional psychometrics revealed strong practitioner understanding of important measurement concepts and principles, particularly regarding validity. Teachers made clear their intuitive understanding of the injunction to use multiple measures in order to make valid inferences and decisions regarding a student's achievement, as specified in the *Standards for Educational and Psychological Testing*:

> Standard 11.20. In educational, clinical, and counseling settings, *a test taker's score should not be interpreted in isolation*; collateral information that may lead to alternative explanations for the examinee's test performance should be considered. (AERA, AEA & NCME, 1999, p. 117, emphasis added).

Similarly, the standards for educational accountability systems developed by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the Consortium for Policy Research in Education (CPRE) prominently and succinctly state, "Decisions about individual students should not be made on the basis of a single test" (Baker, Linn, Herman & Koretz, 2002, p. 3). The American Evaluation Association, in its first public policy pronouncement, has counseled against "simplistic application of single tests or test batteries to make high stakes decisions about individuals and groups [which] impede rather than improve student learning" (2002, unpaginated). The National Association for the Education of Young Children (NAEYC) has issued a position statement declaring:

> Decisions that have a major impact on children such as enrollment, retention, or assignment to remedial or special classes should be based on multiple sources or information and should *never* be based on a single test score. (NAEYC, 1988, emphasis added)

The teachers interviewed spoke of their own multiple measures as providing more accurate portrayals of their students abilities than the state test could provide. The WASL, said one, was merely "one window into a child for one week. As a teacher, I can tell you about their growth as a student" (Ms. Hand). All the teachers agreed that frequent and varied methods were needed to understand and represent accurately the diverse accomplishments of their students.

Washington state relied essentially on the WASL, (Note 6) although the awareness of the importance of multiple measures was indicated in such public statements as the following:

> No single test can tell you everything about a child's performance. Looking at information from a variety of tests and assessment tools remains the best way for parents and classroom teachers to really see how well individual students are learning. (Office of the Superintendent of Public Instruction website, www.k12.wa.us, June 20, 2002)

Construct validity. Some of the teachers interviewed explicitly challenged the WASL's construct validity, questioning whether the test did, in fact, test what it purported to test–*the construct of student achievement*. For example, one teacher said, "There is too much confusion about what it is actually trying to measure" (Ms. Nunn). When scores reflect things besides the intended construct (i.e., *rival constructs*), test results can be misleading, either exaggerating achievement or denying due credit.

The math problem-solving section of the WASL was perceived by some interviewees as troublesome on these grounds, requiring students to explain their solution procedures. Teachers reported that many students who were good at math but weak in writing were unfairly penalized. Said one teacher, "Even if they can explain their thinking and they have the answer right, they get marked down because of their writing skills" (Ms. Hallo). Problems related to rival constructs were not limited to writing requirements in the math test. Teachers suggested several rival constructs actually being measured rather than (or in addition to) the intended construct, student achievement, in saying:

> *Rival construct–socioeconomic status of individual students*: "I'm sure the WASL scores will be best correlated to how much does your mom and dad make economically. . . . Socioeconomic status is the greatest predictor of student success." (Mr. Ming)
>
> *Rival construct–personal difficulties*: "There are other variables that go into testing–a baby, a job, living in their cars. These affect test performance." (Ms. Hand)
>
> *Rival construct–intelligence*: "I firmly believe that WASL performance is not only affected by teaching but also [by] cognitive abilities which, to a certain extent, are innate." (Mr. Exeter)

WASL scores are used not only as measures of the achievements of individual students but "to evaluate instructional practices" and "to hold schools accountable for student learning" (Washington State Senate, 1992, p. 10). For this reason, the validity of inferences on *the construct of school or educational quality* emerged as relevant in the analysis. Several teachers' comments indicated realization that a school's test results might indicate not the quality of its educational program delivery but, rather, the characteristics of its student body, including student motivation and especially affluence:

> *Rival construct–student motivation*: "[Some] students perform badly on the WASL intentionally to make a point." (Mr. Dustin)
>
> *Rival construct–socioeconomic status of school population*: "[A school in my district] traditionally has been at the top but, since we've redistricted, they had a huge influx of students from the lower echelon housing and economic development. That has changed their dynamics. They didn't do as well as they had hoped [on WASL scores]. . . . [S]tudents who are socioeconomically deprived don't do as well." (Ms. Roberts)

Content validity. In describing the test as "much too narrow a device" (Ms. Underwood), one teacher implied that not only construct validity but also content validity was at issue, that the content of the test did not sufficiently represent the content of the intended domain (e.g., the English-language arts test did not fully represenent the domain of English-language arts).

Instructional validity. Relatedly, some teachers indicated that instructional validity–the match between what is taught and what is tested–was faulty, one observing that the WASL "doesn't measure anything that we teach our kids" (Ms. Apple). Another teacher complained that test content was insufficiently aligned with the curriculum:

> I like the fact that people are accountable for teaching certain curriculum, but the assessment part needs work. There is a lot of mismatch between the curriculum and what the WASL is testing. (Ms. Walker)

Scoring concerns

Concerns about the scoring of the state test were also raised. Interpreting a student's written explanation requires professional skill, experience, knowledge of child development, and sometimes knowledge of the particular child, according to one teacher who said:

> The WASL is graded by people with no idea of knowing what good communication is for that child. There's a greater possibility for a disconnect that's unfair for the student. (Ms. Underwood)

The importance of accurate interpretation of text generated by children was not limited to tests of reading and English-language arts. As noted earlier, there were also concerns that students' math achievements might not be fully credited because of the scoring of verbal explanations: "One could be good at math but can't explain their thinking. They would be judged as not passing the test" (Ms. Park).

Two teachers complained that some schools were inappropriately penalized because of regulations related to student scores of zero. One reported that the state had required GED students be classified as sophomores and prohibited them from taking the WASL, then had counted the "lack of scores" from these students against her school, the county's GED school, artificially lowering the school's results (Ms. Apple).

Test expense

A few teachers expressed concern regarding the cost of testing, one preferring a "standardized test which is cheaper and faster" (Ms. Roberts) than Washington's current standards-based (and standardized) test with its performance assessment sections. Another hoped "the state isn't wasting millions of dollars" (Mr. Alder).

Changing state policies and requirements

Some teachers approved the WASL and expected that testing would always be part of the educational system, but one worried about the diversion of resources to the WASL if it proved merely to be "some fad that won't be around long" (Mr. Alder). Another expected no more:

> The WASL is just another one of those things that's going to come, and it, too, shall pass. I haven't changed what I teach or how I teach because, as a conscientious professional, I've looked at what students should know in terms of biology. (Mr. Ming)

In fact, changes to state accountability and testing policy have been enacted "almost every year" (OSPI website, www.k12.wa.us, June 20, 2002) since SSB 5953 in 1992 (see Table 4). Frequent changes, creating layers of increasing and sometimes conflicting requirements, can be seen across the country as state testing programs have increased during the last decade, partly in response to federal requirements regarding Title 1 funding, and in the new federal requirement to test all children in grades 3-8 every year. "Policy hysteria" (Stronach & Maclure, 1996) is a term which has been given to frequent, overlapping policy changes in general (i.e., not necessarily related to testing).

Table 4
Summary of state statutes regarding Washington's education reform initiative

year	bill	effect
1992	SSB 5953	established the framework for education reform and the Commission on Student Learning (expired 1999), providing for the development of the Essential Academic Learning Requirements (EALRs) and a new assessment system.
1993	ESHB 1209	resulting from work by the Governor's Council on Education Reform and Funding (GCERF), established new learning goals and Student Learning Improvement Grants (SLIG) and other programs to help educators help students meet new standards.
1994	ESHB 2850	established requirements pertaining to character traits and values.
1995	SSB 5169	made relatively minor changes to prior law.
1997	ESB 6072	established a timeline for assessment development.
1997	ESHB 2042	established a grade 2 reading assessment.
1998	ESHB 2849	required district school boards to establish reading improvement goals. Also, a grade 4 NRT was moved to grade 3. Also, the legislature provided funds for professional development, instructional materials, and schools with reading programs involving volunteer mentors.
1999	ESHB 5825	made changes to the NRTs and modified the assessment implementation timeline.
1999	SSB 5418	established the Academic Achievement and Accountability Commission, established mathematics goals, and created several new assistance programs.
2002	ESB 6456	authorized the A+ Commission to set performance improvement goals for all students (e.g., economically-disadvantaged students, limited English proficient students, students with disabilities, and students from disproportionately underachieving racial and ethnic backgrounds) and to establish high school graduation rate goals and dropout reduction goals for grades seven through twelve.

Source: Website of the Office of the Superintendent of Public Instruction, State of Washington, www.k12.wa.us, June 20, 2002

Findings

Variations among the perspectives of the 31 teachers interviewed signal continuation of a robust collective struggle to understand and improve education. The variations also evidence the kind of diversity and local control which many have considered traditional strengths of American schooling. The contrasts were so dramatic that two interviewers were "not shocked but stymied" in trying to analyze the range of opinion expressed by the four teachers they had interviewed–perceptions of the state test ranging from approval to ignorance to objection, perceptions expressed with a range of emotions from candor to arrogance to wariness, perceptions varying as to whether the teachers' own assessment practices should follow state mandates or personal beliefs.

Several interviewers expressed surprise that teachers were not more negative about state testing but, instead, that some had offered positive comments or described the state test as a tool to help their teaching. Other interviewers were taken aback by teachers' deep distress about the test and its implications, two interviewers writing, "We feel as overwhelmed as the teachers." Overall, the teachers in this study, like teachers across the country (Shore, 2002), appeared to be adapting and trying to make things work. From the data they provided, four main findings emerged.

(1) *The teachers did not fear accountability but opposed accountability based on a single-shot test.* Their opposition reflected better understanding of the important principle of multiple measures than was manifested in the state accountability policy. Teachers' intuitive, experiential understanding–sometimes referred to as "practical wisdom"–appeared to be stronger in this regard than the formal understanding of state officials and their testing contractors and consultants who had implemented a test-driven accountability system with heavy reliance on the WASL.

(2) *The WASL was not appropriate for children who were eligible for special services, who were non-proficient speakers in English, or who were living in impoverished or marginal situations*, according to teachers who worked with them day-to-day. Teachers indicated that *the state test ensured that these children would not only be left behind but also pressured and punished for factors beyond their control. Individual student scores aggregated and reported as school scores similarly pressured and punished teachers for factors beyond their control*, said some.

(3) Teachers repeatedly claimed *classroom assessments were more informative but sidelined by the state tests.* One teacher, for example, referred to the WASL as "one measurement done during a short period of time that provides a little glimpse of the students, [whereas] I have them all year so I have a better perspective on them" (Ms. Hand). Teachers already understood the message researchers have been trying to share with policymakers, for example:

Once-per-year accountability tests can't do the job of day-to-day, week-to-week pupil diagnosis. . . . What large-scale assessment can't do is document in sufficient detail the what and how of student understandings. (Shepard, 2002)

Policy-makers need to support the development of new assessments and to avoid reliance on single tests. They should shift resources from large-scale assessment to classroom assessment. (Pellegrino, 2002)

(4) *While some teachers appreciated the focus provided by state standards and testing, other teachers were troubled by the test's replacement of teachers' professional judgment*:

The WASL goes against everything we know about learning and takes assessment out of the hands of educators and puts it into the hands of a corporate organization out for profit. (Mr. Twain)

The WASL is robbing me of my professional judgment and replacing learning with inappropriate practices. (Ms. Quinn)

If "inappropriate practices" are the result of state testing, teachers should resist. Although some have blamed teachers' insufficient resistance for the current wave of high-stakes testing (Popham, 2001), some teachers in this study indicated staunch resistance to Washington's state test within their classrooms, their clearest spheres of influence and the location of their primary responsibilities.

High stakes testing represents a mechanism to ensure local compliance to policy initiatives typically described as "reform." Efforts to comply were evident in this study. It nevertheless seems unlikely that centralized, top-down, state control can lead to better education, as implied by the term "reform" (see Fullan, 1991; Sarason, 1990) when it simultaneously deprofessionalizes teachers by usurping their authority and opportunity to plan and implement educational opportunities for their students. In a postmodern era skeptical of grand plans and centralized management, it is worth considering whether forcibly turning teachers into technicians, a return to the previous century's "technological perspective" for controlling education (Hargreaves, Earl & Schmidt, 2002) or "technicist approach" for making education efficient (Gillman, 2002), is more likely to re-form education in a detrimental rather than in an improved manner.

Conclusion

During the data analysis phase of this study, Washington state superintendent Terry Bergeson publicly and plaintively remarked that, as a former school counselor she was not initially an advocate of large-scale testing, but "we need data" (2001). Four months later, a district administrator from Kansas City complained, "We're drowning in data but parched for information–and the questions are cosmic" (Wright, 2002). This study suggests that teachers, who face cosmic questions in the microcosms of their classrooms, are a source of information that policy-makers would be well-advised to heed.

It is no small matter that more than two-thirds of U.S. teachers consider their state tests not worth the investment (Abrams, 2002) and that some teachers are leaving the profession because of test pressure (Gillman, 2002). Teachers are crucial to educational reform not only for the well-known reason that top-down mandates succeed only with bottom-up buy-in from implementers (Fullan, 1991; Sarason, 1990). In addition, teacher perspectives are key to implementing *reasonable* accountability (Shore, 2002) because it is teachers who bring together understanding of children, their achievements, and how to assess them. Teachers' understanding of assessment, despite deprivation of strong formal training, has been too long underestimated. The data clarified a critical difference of emphasis: teachers' focus on "testing *students*" and state or external focus on "*testing* students."

Moreover, understanding teachers' experiences and perspectives helps to explain research findings regarding "perverse incentives" related to state tests, such as teachers' unwillingness to accept or keep positions in low-scoring schools that most need their expertise and energies (Trent, 2002; see also Lankford, Loeb & Wyckoff, 2002). At a time when teacher shortages and high turn-over rates are a matter of concern in Washington state, careful consideration is needed in policy- making circles regarding the impact of test-driven accountability on teacher recruitment, retention, and job satisfaction.

Notes

1. Only 19 states ever reached full compliance (*Education Week*, April 17, 2002, p. 29), and the national system of tests was never developed.
2. Prior to NCLB, NAEP was voluntary for states.
3. The date for making passing the WASL a graduation requirement has been extended to 2008.
4. For permission to use their interiview data and for review of a draft of this manuscript, the authors wish to thank Kevin Crouch, Candace Dawson, Patrick Dowell, Daniel Getty, Jeff Herzog, Stephen Klauer, Karissa Lowe, Jennifer Megli, Mark Muckerheide, Mary Nelson, Wayne Storer, Debra Tidd, and Chad Towe. The authors also thank Marv Alkin of UCLA for review and comments regarding a draft of the article.
5. Since each graduate student interviewed two teachers, there should have been an even number of teachers in the sample. However, by chance and without realizing it, two students chose and interviewed the same teacher.
6. The state also mandated administration of the Iowa Test of Basic Skills (ITBS) in grade 3 reading and math and in sixth grade reading, language arts, and math; and of the Iowa Test of Educational Development (ITED) in grade 9 reading, language arts, math, and an interest inventory. (Source: www.k12.wa.us)

References

Abrams, L. (2002, April). Multi-sate analysis of the effects of state-mandated testing programs on teaching and learning: Results of the national survey of teachers. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.

American Evaluation Association Task Force on High Stakes Testing. (2002). Position statement on high stakes testing in preK-12 education. Fairhaven, MA: AEA.

American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. (1990). *Standards for Teacher Conpetence in Educational Assessment of Students*. Washington, D.C.: Authors.

Baker, E. L. (2002, April). Validity issues for accountability systems. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.

Baker, E. L., Linn, R. L., Hermar, J. L., & Koretz, D. (2002). Standards for educational accountability systems. *CRESST Line*, Winter, 1-4.

Bergeson, T. (2001, December 6). Washington reform update and implications for the future. Presentation to the Washington State Assessment Conference, Seattle, WA.

Bond, L. A., Braskamp, D., van der Ploeg, A., & Roeber, E. (1996). *State student assessment programs database, school year 1994-95*. Oak Brook, IL: Council of Chief State School Officers and the North Central Regional Educational Laboratory.

Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton-Mifflin.

Cannell, J. J. (1987). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice, 7* (2), 5-9.

Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Denzin, N. K. (1997). *Interpretive ethnography: Ethnographic practices for the 21st century.* Thousand Oaks, CA: Sage.

Denzin, N. K. & Lincoln, Y. S. (1994). *Handbook of qualitative research.* Thousand Oaks, CA: Sage.

Education Week (2002, April 17). 1994 ESEA: The state of state compliance. Authors, p. 29.

Elmore, R. (2002, April). Stakes for whom? Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York: Macmillan.

No child left behind (NCLB), reauthorization of the Elementary and Secondary Education Act, Public Law 107-110 (2001).

Fontana, A. & Frey, J. H. (1994). Interviewing: The art of science. In Denzin, N. K. & Lincoln, Y. S. (Eds.), *Handbook of qualitative research* (pp. 361-376). Thousand Oaks, CA: Sage.

Fullan, M. (1991). *The new meaning of educational change.* New York: Teachers College Press.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gewertz, C. (2002, April 10). Low-scoring charter school to shut down in Chicago. *Education Week*, p. 4.

Gillman, C. (2002, April 26). From blooming flowers to marching soldiers: A case study of one kindergarten teacher. Colloquium, Washington State University Vancouver.

GOALS 2000: Educate America Act, Public Law 103-227 (1994).

Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher, 20* (5), 2-7.

Haertel, E. H. (2002, April). Technical considerations in the use of NAEP to confirm states' achievement gains. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives, 8* (41). (http://epaa.asu.edu/epaa/v8n41/)

Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal, 39* (1), 69-95.

Hodder, I. (1994). The interpretation of documents and material culture. In Denzin, N. K. & Lincoln, Y. S. (Eds.), *Handbook of qualitative research* (pp. 403-412). Thousand Oaks, CA: Sage.

Horn, C. (2002, April). Multi-state analysis of the effects of state-mandated testing programs on teaching and learning: Results of the fieldwork studies. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis, 24* (1), 37-62.

LeCompte, M. D. & Preissle, J. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). San Diego: Academic Press.

Lewis, S. (2002, April). What will be the effects on assessment and accountability in local school districts of the "no child left behind" legislation? Presentation to the annual meeting of the National Council of Measurement in Education, New Orleans, LA.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29* (2), 4-16.

Mabry, L. (2002). In living color: Qualitative methods in educational evaluation. In D. Nevo & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation.* Boston: Kluwer-Nijhoff.

Mabry, L. (1999). *Portfolios plus: A critical guide to alternative assessments and portfolios.* Thousand Oaks, CA: Corwin Press.

Mabry, L. (1998). Case study methods. In H. J. Walberg & A. J. Reynolds (Eds.), *Evaluation research for educational productivity* (pp. 155-170). Greenwich, CT: JAI Press.

Mabry, L., Aldarondo, J., & Daytner, K. (1999). Local administration of state-mandated performance assessments: Implications for validity.

Paper presentation to the annual meeting of the American Educational Research Association, Montreal, Canada.

Mabry, L. & Daytner, K. G. (March, 1997). State- mandated performance assessment. Paper presentation to the annual meeting of the American Educational Research Association, Chicago, IL.

Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62* (3), 279-300.

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* San Francisco: Jossey- Bass.

Meyer, L., Orlofsky, G. F., Skinner, R. A., & Spicer, S. (2002). The state of the states. In *Quality counts 2002. Building blocks for success* (a report on education in the 50 states by the Editorial Projects in Education). *Education Week, 21* (17), 68-92.

Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

National Association for the Education of Young Children. (1988). NAEYC position statement on standardized testing of young children 3 through 8 years of age, adopted November 1987. *Young Children, 43* (3), 42-47.

Neuman, S. (2002, April). The Bush accountability and assessment agenda: New opportunities and challenges. Presentation to the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Orlofsky, G. F. & Olson, L. (2001). The state of the states. In *Quality counts 2001: A better balance* (a report on education in the 50 states by the 2001 Editorial Projects in Education). *Education Week, 20* (17), 86-92, 94-100, 102- 106.

Oregonian. (2002, November 29). Nearly all Washington schools fail to hit goals. Portland, OR: Authors, p. B2.

Pellegrino, J. (2002, April). Assessment and learning: Issues highlighted in the NRC report "Knowing what students know." Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Pomplun, M. & Capps, L. (1999). Gender differences for constructed-response mathematics items. *Educational and Psychological Measurement, 59* (4), 597-614.

Popham, W. J. (2001). *The truth about testing: An educator's call to action.* Alexandria, VA: Association for Supervision and Curriculum Development.

Preusch, M. (2001, December 15). National briefing Northwest: Oregon unemployment rate rises. *New York Times*, p. 14.

Rubin, H. J. & Rubin, I. S. (1995). *Qualitative interviewing: The art of hearing data.* Thousand Oaks, CA: Sage.

Sarason, S. B. (1990). *The predictable failure of educational reform: Can we change course before it's too late*? San Francisco: Jossey-Bass.

Schoenfeld, A. (2002, April). "This is just a test!" Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Shepard, L. A. (2002, April). Building bridges between classroom and large-scale assessments. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Shepard, L. A. & Smith, M. L. (1988). Escalating academic demand in kindergarten: Counterproductive policies. *Elementary School Journal, 89* (2), 135-145.

Shore, A. (2002, April). Optimizing the validity and value in the public debate over testing as a tool in educational reform. Paper presentation to the annual meeting of the American Educational Research Association, New Orleans, LA.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20* (5), 8-11.

Smith, M. L. & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice, 10* (4), 7-11.

Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher, 7* (2): 5-8.

Sternberg, R. J. (2002). The "Janus principle" in psychometric testing: The example of the upcoming SAT-I. *The Score*, newsletter of American Psychological Association Division 5, Evaluation, Measurement, and Statistics, *24* (2), 3–5.

Stiggins, R. J. & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* Albany, NY: SUNY Press.

Stronach, I. & Maclure, M. (1996). Mobilizing meaning, demobilizing critique? Dilemmas in the deconstruction of educational discourse. In *Cultural Studies* (vol. 1, pp. 259- 276). Greenwich, CT: JAI Press.

Swanson, C. B. & Stevenson, D. L. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation and Policy Analysis, 24* (1), 1-27.

Trent, W. (2002, April). The policy implications of federally mandated annual testing. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

U. S. Census Bureau. (2001). Percent of people in poverty by state. Online at http://www.census.gov/prod/2001pubs/p60-214.pdf.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process.* Cambridge, MA: Harvard University Press.

Washington State Senate. (1992). Washington Substitute Senate Bill 5953: Act relating to education. Olympia, WA: Authors.

Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation.* Thousand Oaks, CA: Sage.

Wright, D. D. (2002, April). Who did we miss, and why? Factors associated with non-participation of general education students in standardized assessments. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Yakimowski, M. (2002, April). What will be the effects on assessment and accountability in local school districts of the "no child left behind" legislation? Presentation to the annual meeting of the National Council of Measurement in Education, New Orleans, LA.

About the Authors

Linda Mabry is an associate professor at Washington State University Vancouver, where she specializes in assessment of student achievement, program evaluation, and qualitative research methodoloby, and a member of the boards of the American Evaluation Association and the National Center for the Improvement of Educational Assessment.

Jayne Poole is a graduate student in Education at Washington State University Vancouver, where she is researching the reading- writing connection, and a kindergarten teacher of eleven years in Longview, Washington.

Linda Redmond, a recent Masters in Education graduate of Washington State University Vancouver, has taught in the public schools of Washington state for twenty-two years and is currently an elementary music specialist in Longview, Washington.

Angelia Schultz, a newly certificated teacher with a BA in English and a graduate student in Education at Washington State University Vancouver, currently works as a substitute teacher. Her interests include the consequences of high stakes testing and test- driven accountability.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós

Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 1077 times since July 31, 2003

Volume 11 Number 23 **July 31, 2003** **ISSN 1068-2341**

A National Crisis or Localized Problems?
Getting Perspective on the Scope and Scale of the Teacher Shortage

Patrick Murphy
University of San Francisco

Michael DeArmond
Kacey Guin
University of Washington

Citation: Murphy, P., DeArmond, M., Guin, K. (July 31, 2003). A National Crisis or Localized Problems? Getting Perspective on the Scope and Scale of the Teacher Shortage. *Education Policy Analysis Archives, 11*(23). Retrieved [date] from http://epaa.asu.edu/epaa/v11n23/.

Abstract

Despite the considerable attention the popular press has devoted to the question of teacher shortages, there have been surprisingly few attempts to systematically measure the size and nature of the problem. This article attempts to estimate the size and nature of the celebrated teacher shortage of the late 1990s by using data from the U.S. Department of Education's 1999-00 School and Staffing Survey. While limitations of the SASS data do not allow us to directly estimate the absolute size of the shortage, they do allow us investigate its relative impact. An examination of the data shows that the problem was distributed unevenly: urban schools and those with relatively high populations of minority and low-income students bore the brunt of the shortage; southern and western states had more problems filling teaching slots than other regions did. These findings suggest that state and local officials should keep distributional concerns in mind when they design policies to improve teacher recruitment and retention.

Introduction

Beginning in 1999, concerns about the supply of teachers for the nation's elementary and secondary schools found their way on to the education policy agenda. The headlines at the time told the story: "Help Wanted: 2 Million Teachers," (Note 3) "Districts Step up Teacher Recruiting," (Note 4) "New Teachers are Hot Commodity." (Note 5) The problem appeared to be a crisis. It raised serious questions about the how schools recruited and hired teachers and whether or not they could meet the challenge of the shortage.

As the economy slipped into recession in the fall of 2001, the impending sense of disaster subsided. News coverage noted that the gap was filling in, and the doom and gloom predictions began to recede. (Note 6) Some observers even began to question whether the situation was ever as critical as the public was led to believe. Was there anything to worry about after all?

The question was understandable. At the peak of the reporting on the shortage, there was surprisingly little systematic information about the problem's impact. Policy-relevant data to guide state and local decision makers was particularly hard to find. Instead, the issue was often presented in a dramatic, one-two punch: an anecdote about a school or district struggling to hire teachers followed by dire statistical warnings about the problem's overwhelming national scale. At the time, this impressionistic view of the problem left some decision makers to speculate about the shortage's effect while reviewing their policy options. Today, the residue of this anecdotal formulation makes it hard to pin down what actually happened.

And yet, understanding what actually happened can offer insight into an enduring and critical question: how can we provide an adequate supply of quality teachers for our schools? The importance of this question clearly remains, regardless of the health of the nation's economy.

With that in mind, this article attempts to offer some insight into the reported shortage of teachers that the country recently experienced and to disaggregate its impact. Our general goal is to break down what was characterized as an amorphous and somewhat monolithic issue – *the teacher shortage* – into more meaningful terms for public policy decision-making. We attempt to understand the problem's scale and scope by using the recently released National Center for Education Statistics (NCES) 1999-00 School and Staffing (SASS) data.

This article is arranged as follows. In Section II we provide background on the supply and demand of teachers for public schools. Section III reports findings from our analysis of the SASS data. Here we measure the impact of the teacher shortage across geographic, socioeconomic, and other dimensions. In Section IV we present what our findings imply for public policy. We conclude that, from a national perspective, it is not very useful to speak of a monolithic teacher shortage. Instead, the SASS data support what many observers have contended for some time: the impact of the shortage is unevenly distributed across schools and school districts. This unevenness suggests that policies designed to improve the recruitment and training of teachers need to focus on distribution issues as well as questions of quantity and quality.

II. Background: teacher supply and demand

Size of the Shortage

Despite considerable media attention, there have been relatively few attempts to quantify a national teacher shortage. One of the more frequently cited figures about the problem suggests that the nation will need to hire 2.2 million teachers over the next decade. This number can generally be traced to Hussar (1999) who, using the same model for three different scenarios, predicted the nation's districts would need to hire between 1.7 million and 2.7 million teachers between 1998-99 and 2008-09. Hussar's model accounted for predicted growth in the student population and included varying levels of teacher continuation, pupil/teacher ratios, and teacher age distributions. It assumed relatively modest increases in the demand for teachers, placing growth rates between 1 and 4 percent per year. A separate analysis by Wayne (2000) produced similar estimates. That analysis concluded that there would be a two to three percent increase per year in the number of teachers needed over the next decade. Both Hussar and Wayne suggest there is and will be a national need of approximately 200,000 new teachers each year for the foreseeable future.

Researchers have conducted similar projections for individual states. The Illinois State Board of Education, for example, estimated that the state would need 60,000 new teachers in the next three years (Banchero and Spencer, 2000). Perry (2001) estimated that California's schools would need as many as 300,000 new teachers over the next ten years.

These figures cannot help but grab the public's attention. They are large, and they conjure up daunting images. Unfortunately, they do not tell us anything about the size or distribution of a possible *shortage*. They merely estimate the future demand for teachers without relating it to any projections regarding teacher supply.

State-level analyses appear to be the only ones to provide both sides of the equation with a few estimating both teacher demand and supply. Estimates from the Florida Education Department, for example, suggested that the state would need about 12,000 more teachers per year *than are projected to be supplied* (Office of Strategy Planning, 2000). If correct, this would be a dramatic situation: over 8 percent of the state's teaching positions could go unfilled. In North Carolina, state education researchers calculated a shortage of more modest proportions, estimating that current demand will outstrip supply by about 2,000 teachers over the next decade. This is less than two percent of the state's elementary and secondary teacher population (NC State Department of Public Instruction, 1998).

Causes of Teacher Shortage

Though national shortage estimates are hard to find, there are many attempts to explain what drives the demand for teachers. At the top of the list of contributing factors is teacher turnover. Given the aging of the teaching force -- the average age of teachers has increased steadily over the past 10 years (Hussar, 1999) -- some see future demand as being driven by a wave of retirees. Others point to pre-retirement attrition. U.S. Education Department statistics, for example, suggest that as many as 9 percent of new teachers quit during their first year of teaching and as many as one-out-of-five teachers leave in the first three years (Yasin, 1999). Some of these teachers leave permanently to pursue a different career; others leave temporarily. These temporary leavers represent significant numbers. Nationally, one-quarter of the teachers hired each year are people who, though not currently teaching, have some prior teaching experience (Wayne, 2000).

There is considerable debate over which factors are behind this pre-retirement attrition. Ingersoll (2001) argues that organizational factors within a school -- low salaries, lack of support from administrators, student discipline issues, and lack of input and decision-making power -- cause teachers to leave their position (or the teaching profession altogether). Harrington (2001) blames the specific shortage of math, science and technology teachers on "a dysfunctional labor market held hostage by poor allocation of resources, disincentives to productivity and, ironically, inequity" (2001: 8). Equal pay for all teachers, he argues, distorts the market for teachers in these technical subject areas. Wayne (2000) maintains that people are more apt to leave teaching for family and personal reasons than because they are dissatisfied with their job. With all of this attention to turnover, it is easy to forget that, when compared to other professions, teaching remains one of the most stable employment choices a recent college graduate can make (Henke, Zahn, and Carroll, 2001).

Besides turnover, two other major factors are behind the increasing demand for teachers. In some regions of the country, districts clearly need to hire more teachers to keep up with growing enrollments. Despite this, when we consider the national picture, enrollment growth does not appear to be a big driver of teacher demand. The nation's public elementary and secondary school enrollment, for example, is predicted to increase by only one percent between 1999-2000 and 2010-11; between 1988-89 and 1999-2000, it increased seventeen percent (Hussar, 2002). The other factor behind teacher demand is class size reduction policies, though it is a phenomenon concentrated in particular states. It comes as no surprise that when states mandate smaller classes, districts need more teachers. In the end, it appears that class-size reduction policies do more to drive the demand for teachers than population growth (Harrington, 2001; Hussar, 1999; Shield et al 2001).

Quantity vs. quality

Despite all of this attention to quantifying teacher demand, many researchers argue that *quality*, not quantity, should be the central focus of any teacher supply discussion. With such a focus, the research is forced to take a more complex look at supply and demand in the teacher labor market (Broughman and Rollefson, 2000).

Barker and Smith (1997), for example, note that the percentage of teachers teaching out of field (i.e. those not holding a major or minor degree in the subject that they teach) is on the rise. This finding suggests there is a teacher quality shortage rather than a teacher quantity shortage. Ingersoll (1997) echoes this sentiment, stating that while many schools report difficulty in finding quality teachers, few have problems just filling positions. This argument for quality, however, assumes that there is widespread agreement on what constitutes a quality teacher (i.e., certification, major in subject area, etc.). Unfortunately, there is at present no such agreement. (Note 7)

Distribution of Teachers

In addition to teacher demand and teacher quality, researchers have looked at the distribution of teachers among different kinds of schools and districts, suggesting that quantity and quality vary across subject areas, geographic regions, and social and economic dimensions.

Math, science, and foreign language teachers often lead the list of high demand subject areas. A Texas study of teacher supply and demand for the 2001-02 school year, for example, found that most districts in the state were able to hire enough teachers to fill their vacancies. The problem areas were secondary level teachers in four subjects: science (3 percent unfilled at the start of the school year), foreign language (16 percent), technology (10 percent), and bilingual/ESL (26 percent) (Sparks, 2002).

Though it did not emerge in the Texas study, special education is also considered a high demand subject area across the nation (Hare, Nathan, and Darland, 2000; Sack, 1999; The Urban Teacher Challenge, 2000). In the case of special education, there is evidence of a combined quality and absolute quantity shortage. Boe, et al (1998), for example, found that the percent of special education teachers who lacked full certification ranged from 8-10% in the years between 1984-85 and 1992-93. This was almost twice the percentage of regular education teachers who lacked full certification. Given that the number of children identified for special education has risen over the past 10 years, it would appear likely that that percentage has continued to increase. (Note 8)

Other research has indicated that the impact of quality shortages is distributed unevenly across location and social class. Shields et al (2001), for example, found that the bulk of teacher shortages in California were concentrated in urban, low income, low performing, and minority schools. During the 2000-2001 school year, urban schools had on average 19% uncertified teachers, compared with 9% in suburban and rural schools. Carroll and his colleagues reached similar conclusions about the distribution of teachers in California. They noted that when teachers moved from one district to another, or from school to school within a district, they were likely to move to schools that served fewer minority students and fewer students eligible for free and reduced lunch programs (Carroll, et al, 2000).

Finally, school districts appear to be finding it increasingly difficult to assemble a diverse group of teachers for their schools. Several different sources have identified a shortage of teachers of color as another dimension of the teacher supply question (Grissmer and Kirby, 1997; Kirby, Berends and Naftel, 1999; Lewis, 1996; The Urban Teacher Challenge, 2000).

The literature discussed above primarily focuses on the demand for teachers. While some information is available about teacher quality and distribution, adding valuable perspective on shortages of certified teachers, those studies are generally limited to state-level data. Finally, it is unclear from any of the research how many districts start the year with unfilled teaching positions and which students those districts serve. The next section uses national data to examine the impact of teacher shortages on district efforts to fill open positions. While data limitations preclude a precise estimate of the shortage, the analysis does take into account distributional issues by measuring the impact of the shortage across different social and geographic dimensions.

III. Calculating Estimates

Over the last 15 years, the U.S. Education Department's National Center for Education Statistics (NCES) has used its Schools and Staffing Survey (SASS) to collect information on staffing and personnel issues in the nation's K-12 schools. NCES's most recent effort, the 1999-2000 SASS, involved a sample of public schools, district offices, teachers, principals, as well as public charter schools. Private schools and Bureau of Indian Affairs schools also participated in the survey. NCES selected the respondents so as to provide a nationally representative database of public K-12 teachers, principals, schools, and school districts (U.S. Department of Education, NCES 2000: 2). Collectively, the survey questions covered a wide range of issues, including: school and district capacity, descriptive demographics, teacher training and experience, salary structures, instructional practices, parent involvement, and the use of technology.

The 1999-00 SASS cycle also included new questions designed to provide information about different aspects of teacher supply and demand (NCES, 2000a, p.3). Though these items fall far short of providing estimates of the size of the shortage, the survey includes two areas of inquiry that can shed some light on these issues. Together they provide a useful backdrop for our work.

First, the SASS asked school districts about how many total teachers they employed and about the timing of their new hires. It is possible to use these questions to estimate the relative share of total teachers that were hired after the start of the school year. This *late-fill rate* provides one, albeit imperfect, (Note 9) indicator of teacher shortages during the 1999-2000 school year across different districts. Second, items in the school questionnaire attempted to assess how hard it was for schools to hire teachers for particular subject areas. Together, these portions of the survey can provide a more systematic, if qualified, picture of the shortage compared to anecdotal accounts found in the media.

Findings: The Scale and Impact of the Shortage

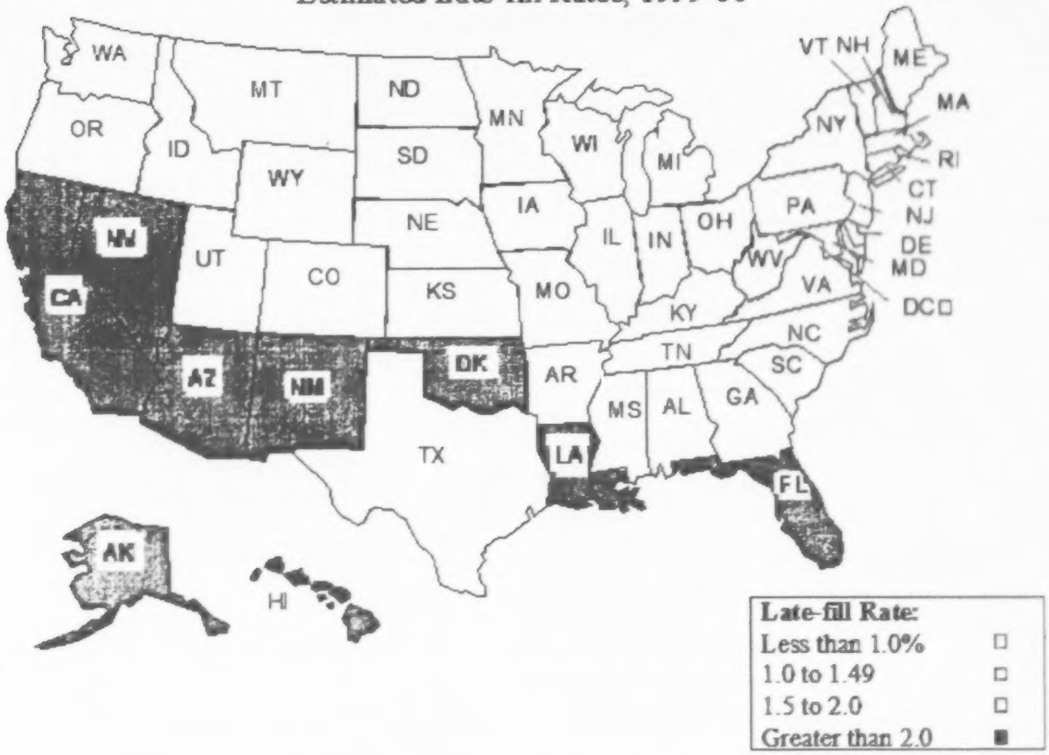
Using the SASS data, an estimated 45,000 (Note 10) were hired after the start of the 1999-2000 school year suggesting that at least these many positions were unfilled in public schools when school began. (Note 11) This figure represents 1.5 percent of the total teaching positions in public schools (based on a national estimate of 3 million). Because the SASS tells us only about teaching positions that were eventually filled, this does not capture the number of positions that were never filled. As such, the 45,000 number understates the total number of vacancies. Indeed, given the limitations of the SASS data it is impossible to estimate an *absolute* vacancy rate for districts.

Nevertheless, if we assume that a district's late-fill rate generally reflects the overall vacancy rates in its schools, we can use late-fill rates to examine *relative* variations in the shortage problem across districts. With this in mind, a further look at the data show that the impact of the problem is not distributed evenly. (Note 12)

Regional Distribution

A regional analysis of the SASS data supports the conventional wisdom that some regions of the country have more to worry about with regard to teacher shortages than others. Using a late-fill rate estimate calculated from the SASS data, we were able to create state level late-fill estimates. Figure 1 shows how those late-fill estimates vary across the country. (Note 13) As the figure suggests, several states significantly exceed the national average (1.5 percent). Among the highest are five states with Hawaii (5.9 percent) and Alaska (5.6) leading the list, followed by New Mexico (2.6), Arizona (2.4), and California (2.3). States in parts of the southeast also emerge with relatively high late-fill rates.

Figure 1
Estimated Late-fill Rates, 1999-00



Midwestern states, by contrast, appear to have less difficulty in hiring teachers. Most of these states, covering a band from Pennsylvania in the east to Idaho in the west, had a late-fill rate of less than 1.0 percent. Iowa represents the limit case, with an estimated late-fill rate of 0.4 percent at the start of the year. Given some minimal amount of personnel shifts at the last-minute, one would expect a certain number of positions to be filled after the start of the school year. The 0.4 percent figure, then, could be considered very close to a zero rate of vacancies.

Subject Field Variation

Beyond the regional variation, observers also have, as noted above, suggested that the need for teachers varies across different subject areas. An examination of the data supports this idea. Although the SASS data does not lend itself to a *late-fill rate* analysis by subject area, the school questionnaire did differentiate by subject area when it asked schools how difficult it was to fill particular positions. Because the survey response options were qualitative (respondents could choose from options like "easy…somewhat difficult…difficult…") the results, especially those involving comparisons, should be interpreted with caution. (Note 14) Nevertheless, it is possible to identify which subject areas were generally perceived by schools as being the hardest to fill.

Table 1 presents calculated national estimates of the average *difficulty score* schools reported for different subject areas. Special education, foreign language, and English as a second language top the list. Positions in math and the physical sciences were also difficult to fill. Interestingly, schools reported that vocational education instructors, a subject area that does not get much attention in either media reports or academic research, were just as hard to find as special education teachers. At the other end of the spectrum, public schools found it relatively easy to find English, social studies, and elementary school teachers.

Table 1. Difficulty in Hiring of Different Subjects as Reported by Schools*

Subject	*Avg. Difficulty Score*
Foreign Language	2.28
Special Education	2.19

Vocational Education	2.19
ESL	2.11
Math	2.10
Physical Sciences	2.03
Computer Science	1.99
Biology	1.95
Music/art	1.90
English	1.55
General Elementary	1.39
Social Studies	1.36

*Where 1=easy, 2=somewhat difficult, 3=very difficult, 4=position never filled.

Source: Calculated from 1999-00 School and Staffing Survey, National Center for Education Statistics, U.S. Department of Education.

A Disproportionate Impact

The literature discussed above also suggests that, in addition to the variation by region and subject, other factors have an effect on how difficult it is for a school or district to fill a teaching position. Using the same definition of unfilled teaching positions discussed above, Table 2 offers estimates of the late-fill rate at the start of the school year for districts in urban, suburban, and rural areas, those with a high percentage of minority students, and the relative percentage of students eligible for free and reduced lunch programs. The data suggest that the challenge of hiring teachers becomes less difficult as one moves away from the central city. The late-fill rate for urban school districts was more than 50 percent higher than that for suburban school districts and twice as high as the figure for rural schools. Another way of looking at the disproportionate impact of the teacher shortage on central city schools is to note that though urban districts account for 29 percent of the teaching positions in the country, they represented 41 percent of the late-fill positions for the 1999-00 school year.

Given these figures for urban schools, it comes as little surprise that the shortage has a more profound impact on schools with relatively high minority student populations and larger shares of students eligible for free and reduced lunch. Table 2 separates public school districts into two categories, those whose student population is comprised of more than 40 percent minority students and those districts where minority students account for 40 percent or less of the population. School districts with high minority populations appeared to have a much more difficult time filling their teaching positions in 1999-00. These districts accounted for less than half (42 percent) of the total teaching positions, but they represented over 57 percent of the total number of late-fill positions at the start of the year. That figure translates into a 2.11 percent late-fill rate, or twice the rate of districts with fewer than 40% minority students.

The findings regarding the minority student population are very similar to those that emerge when one examines the impact of the shortage relative to the socio-economic status of the student population. Using the percentage of students eligible for free and reduced lunch as a proxy, districts with relatively high levels of low-income students found a larger share (1.88 percent) of their teaching positions filled after the start of the year compared to those with fewer low-income students (1.13 percent).

**Table 2. Late-fill Rate by Share of Minority Students, Free/Reduced Lunch Eligible and Location of School Districts
SASS 1999-00**

	Total Positions	Unfilled Positions	Late-fill Rate (%)
Percentage Minority Students			
40% or less	1,865,090	19,580	1.05
Greater than 40%	1,207,639	25,508	2.11
Students Eligible for Free/Reduced Lunch			

40% or less	1,693,096	19,211	1.13
Greater than 40%	1,379,633	25,877	1.88
District Location			
Central City	893,067	18,602	2.08
Suburban	1,508,341	20,322	1.35
Rural	671,321	6,163	0.92
U.S. Total	3,072,729	45,088	1.47

Source: Calculated from 1999-00 School and Staffing Survey, National Center for Education Statistics, U.S. Department of Education.

IV. Policy Considerations

In sum, rather than a national crisis, the SASS data show the teacher shortage to be a regional, subject specific phenomenon. Districts in the southeast, southwest, and west had a more difficult time filling vacancies than those in the Midwest and northeast. Foreign language and special education teachers were among the hardest to find, as were vocational education, math, and science teachers. Most importantly, the impact of the shortage is far more acute in lower income, urban districts with relatively high minority student populations. These findings support many of the characteristics of the shortage that researchers found and/or asserted.

While this research adds a national perspective on the problem and a systematic investigation into its various facets, it also illuminates a significant gap that remains in the teacher supply/demand data. It is still not possible to estimate the absolute magnitude of the shortage with any confidence. This problem persists despite efforts by the NCES to add recruitment questions to the most recent survey on school staffing. Though it is helpful to be able to identify the types of schools experiencing the greatest difficulty in hiring, as well as the subject areas in greatest demand, it is still not possible to provide vacancy rate estimates. If policy makers are to design efficient responses to teacher supply problems in the future, information on the scale of the problem is necessary.

From a public policy perspective, the fact that the impact of the teacher shortage problem was so unevenly distributed across schools and school districts suggests a need for targeted efforts to address the problem(s). Policies designed to simply increase the supply of teachers across the board, for example, may shrink the absolute size of a shortage but do little to reduce the relative impact of the problem on poor, urban districts with high populations of minority students. Signing bonuses are an example of such an approach. In the spring of 2001, Governor Jeb Bush proposed a $1,000 signing bonus for all new Florida teachers. Under the plan, a new elementary teacher in a middle-class, predominately white, suburban school would get the same bonus as someone interested in teaching math to poor, minority students in the inner city. Eager to respond to a looming crisis, the legislature passed the blanket bonus even though it probably would do little to address the schools that were truly struggling to find teachers.

Signing bonuses are not the only example of well-intended efforts that will have little impact on the most important aspects of this problem. Many states have recently invested in expanding teacher training programs as well as accelerating their credentialing procedures. While increasing the number of teachers in the pipeline may eventually address the areas of greatest need, such policies could be more focused. A new program that makes it easier for an individual to prepare to teach social studies at a suburban high school will have little impact. Targeting incentives and support for individuals interesting in obtaining special education, math, or science certifications, however, would be a better use of scarce public resources.

The uneven impact of the teacher shortage, therefore, suggests that policy makers need to be more strategic in their response. Programs designed to increase the supply of teachers as well as provide incentives for them to enter the most challenging schools might be more likely to provide assistance for struggling schools. In short, decision makers need to consider the distributional impact of different human resource policy options.

Finally, it is important to note that the issue of how teachers are distributed across schools remains relevant even in the absence of a teacher shortage. Recent state budget crises – and subsequent cuts in education funding -- appeared to have rendered discussions about teacher recruitment moot. While schools may be recruiting fewer teachers overall, the current situation is likely to bring about a significant shuffling of human resources. Just as the impact of the reported teacher shortage appeared to be uneven, one expects this new crisis to disproportionately hit the schools with the greatest needs. As states and districts wrestle with difficult choices about resource distribution, they should take into account how their choices will affect the distribution of teachers across their schools.

References

Baker, D.P. and Smith, T. (1997). Trend 2: Teacher turnover and teacher quality. Refocusing the issue. Teachers College Record. 99:1. pp 29-35.

Banchero, Stephanie and LeAnn Spencer (2000). "Teacher Shortage Looms, Study Warns," Chicago Tribune, December 14, 2000

Boe, E.E., Bobbitt, S.A. and Cook, L.H. (1997). Whither didst thou go? Retention, reassignment, migration and attrition of special and general education teacher from a national perspective. The Journal of Special Education. 30(4). pp.371-389

Boe, E.E., Cook, L.H., Bobbitt, S.A. and Terhanain G. (1998). The shortage of fully certified teachers in special and general education. Teacher Education and Special Education. 21:1. pp.1-21.

Bradley, A. (10 March 1999). States' uneven teacher supply complicates staffing of schools. Education Week.

Bradley, A. (7 April 1999). Crackdowns on emergency licenses begin as teacher shortage looms. Education Week.

Broughman, S.P. and Rollefson, M.R. (2000). Teacher supply in the United States: Sources of newly hired teachers in public and private schools. 1987-88 to 1993-94. (NCES 2000-309). U.S. Department of Education, National Center for Education Statistics. Washington, DC. U.S. Government Printing Office.

Carroll, Stephen, Robert Reichardt, Cassandra Guarino, and Andrea Mejia (2000). The Distribution of Teachers Among California's School Districts and Schools, Santa Monica: RAND, 2000

Darling-Hammond, L. (2001). The research and rhetoric on teacher certification. A response to "Teacher Certification Reconsidered" National Commission on Teaching and America's Future.

Grissmer, David and Kirby, Sheila N. (1997). Teacher turnover and teacher quality. Teachers College Record. 99:1. pp. 45-56.

Hare, D., Nathan, J., Darland, J. and Laine, S.W.M. (2000). Teacher shortages in the Midwest: Current trends and future issues. Minneapolis, MN: Center for School Change. University of Minnesota.

Harrington, P.E. (2001). Attracting new teachers requires changing old rules. The College Board Review. 192. pp. 6-11.

Henke, R.R., Zahn, L. and Carroll, C.D. (2001). Attrition of new teachers among recent college graduates: Comparing occupational stability among 1992-93 graduates who taught and those who worked in other occupations. (NCES 2001-189). U.S. Department of Education, National Center for Education Statistics. Washington, DC. U.S. Government Printing Office.

Hussar, W.J. (1999). Predicting the need for newly hired teachers in the United States to 2008-09. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hussar, W.J. (2002). Projections of Education Statistics to 2011. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Ingersoll, R.M. (1997). Teacher turnover and teacher quality: The recurring myth of teacher shortages. Teachers College Record. 99:1. pp. 41-44.

Ingersoll, R.M. (2001). Teacher turnover, teacher shortages, and the organization of schools. Center for the Study of Teaching and Policy. University of Washington.

Kirby, S.N., Berends, M. and Naftel, S. (1999). Supply and demand of minority teachers in Texas: Problems and prospects. Educational Evaluation and Policy Analysis. 21:1 pp.47-66.

Lewis, M.S. (1996). Supply and demand of teachers of color. ERIC Digest. ERIC Clearinghouse on Teaching and Teacher Education, Washington, DC.

North Carolina State Department of Public Instruction, Department of Human Resources (1998). Teacher Supply and Demand Study.

Office of Strategy Planning (2000). "Projected Number of Teachers Needed Through 2020-2021." Tallahassee: Florida Department of Education (December 2000).

Perry, Mary, (2001). "Update on California Teacher Workforce Issues," Palo Alto, California: EdSource, Inc.

Sack, J.L. (24 March 1999). All class of spec. ed. teachers in demand throughout nation. Education Week.

Shields, P.M., Humphrey, D.C., Wechsler, M.E., Riehl, L.M., Tiffany-Morales, J., Woodworth, K., Young, V.M. and Price, T. (2001). The status of the teaching profession 2001. Santa Cruz, CA: The Center for the Future of Teaching and Learning.

Sparks, K. (2002). *Teacher demand study 2001-2002*. Institute for School-University Partnerships. Bryon, TX

Texas Education Agency (1995). *Policy Research Report #6: Texas teacher retention, mobility, and attrition*. Austin, TX: Texas Education Agency.

U.S. Department of Education National Center for Education Statistics (2000). *Schools and Staffing Survey and Private School Survey Questionnaires, 1999-2000*, Washington, D.C.: U.S. Department of Education (NCES 2000-310) 2000.

The Urban Teacher Challenge (2000). *Teacher demand and supply in the great city schools*. www.cgcs.org and www.rnt.org. Belmont, MA.

Walsh, K. (2001). Teacher certification reconsidered: Stumbling for quality. The Abell Foundation. Baltimore, MD.

Wayne, A.J. (2000). Teacher supply and demand: Surprises from primary research. *Education Policy Analysis Archives*. 8(47). http://epaa.asu.edu/epaa/v8n47.html

Yasin, S. (1999). The supply and demand of elementary and secondary school teachers in the United States. ERIC Digest. ERIC Clearinghouse on Teaching and Teacher Education, Washington, DC.

About the Authors

Patrick Murphy
Associate Professor
Chair
Department of Politics
University of San Francisco
415.422.5867
Email: murphyp@usfca.edu

Patrick Murphy is an Associate Professor of Politics at the University of San Francisco.

Michael DeArmond and Kacey Guin are policy researchers at the Center on Reinventing Public Education, Daniel J. Evans School of Public Affairs, University of Washington.

Appendix

Estimates of Teacher Late-fill Rates by State
SASS 1999-2000

State	Late Hires*	Total Teachers**	Late-fill Rate***
Alabama	922	51,891	1.8%
Alaska	518	9,286	5.6%
Arizona	1,136	47,295	2.4%
Arkansas	286	33,500	0.9%
California	6,896	299,836	2.3%
Colorado	549	44,420	1.2%
Connecticut	646	44,166	1.5%
Delaware	160	8,009	2.0%
District of Columbia	88	5,395	1.6%
Florida	2,974	141,651	2.1%
Georgia	1,831	96,246	1.9%
Hawaii	636	10,735	5.9%
Idaho	70	14,899	0.5%
Illinois	1,977	130,056	1.5%
Indiana	468	61,152	0.8%
Iowa	158	37,823	0.4%
Kansas	257	34,268	0.7%
Kentucky	624	43,341	1.4%
Louisiana	1,216	54,333	2.2%
Maine	161	19,108	0.8%
Maryland	483	51,734	0.9%
Massachusetts	722	80,647	0.9%
Michigan	841	100,752	0.8%
Minnesota	579	63,873	0.9%
Mississippi	551	33,661	1.6%
Missouri	351	66,744	0.5%

Montana	112	11,004	1.0%
Nebraska	149	20,619	0.7%
Nevada	391	19,334	2.0%
New Hampshire	156	16,170	1.0%
New Jersey	1,557	108,809	1.4%
New Mexico	535	20,488	2.6%
New York	3,720	211,724	1.8%
North Carolina	1,509	84,125	1.8%
North Dakota	37	7,878	0.5%
Ohio	649	120,839	0.5%
Oklahoma	904	45,180	2.0%
Oregon	355	31,193	1.1%
Pennsylvania	754	120,522	0.6%
Rhode Island	200	12,899	1.6%
South Carolina	833	46,195	1.8%
South Dakota	67	11,040	0.6%
Tennessee	809	59,317	1.4%
Texas	3,833	266,083	1.4%
Utah	164	23,119	0.7%
Vermont	150	8,885	1.7%
Virginia	1,486	90,181	1.6%
Washington	826	61,943	1.3%
West Virginia	370	20,977	1.8%
Wisconsin	383	61,816	0.6%
Wyoming	39	7,568	0.5%
Total	45,088	3,072,729	1.5%

*Those teachers hired after the start of the school year.

** Teaching positions ("head counts") NOT full-time equivalents.

***Late-fill rate does not include those teaching positions which were never filled.

Notes

1. This research is part of a larger project on the teacher shortage and how districts have responded. Funding for the project was provided by the Smith Richardson Foundation. The views and opinions expressed here are those of the authors.
2. Patrick Murphy is an Associate Professor of Politics at the University of San Francisco. Michael DeArmond and Kacey Guin are policy researchers at the Center on Reinventing Public Education, Daniel J. Evans School of Public Affairs, University of Washington.
3. *Education Week*, Special series, March/April 1999 (http://www.edweek.org/sreports/help.htm)
4. Houtz, J. (21 April 1999) "Districts Step Up Teacher Recruiting" *Seattle Times*.
5. Bradley, A. (9 September 1998) "New Teachers are Hot Commodity." *Education Week*
6. (*Ed Week* article)
7. See Hayasaki, E. (10 February, 2003) "Teacher Shortage Abates" *Los Angeles Times*. Page 1.
8. See Darling-Hammond (2001) and Walsh (2001) for differing perspectives on this issue.
9. Boe et.al.(1998) attributes this ongoing shortage of special education teachers to not enough teachers certified to teach special education are entering the field each year and a high turnover rate. Boe, Bobbitt and Cook (1997) note that special education attrition rates were not significantly different than general education attrition rates, indicating that the high turnover rate result from special education teachers switching to regular education, as opposed to leaving the teaching profession all together.
10. Given the wording and structure of the questions, it is not possible to determine precisely how many vacancies a district may have had at the start of the year. The District Questionnaire first asks how teachers were "newly hired" for the 1999-2000 school year. It then asks, of those new hires, how many were hired before the summer break, during the first half of the summer, etc. The final question in this series asks how many were hired after the beginning of the school year. It is this figure that serves as the numerator for the late-fill rate estimate. The reported total number of teachers in the district (head counts, NOT FTE) is the denominator. Not captured in these figures are those teaching positions that went unfilled during the school year. Therefore, the late-fill rate provides a relative measure of the depth of the shortage *but it is likely to understate in absolute terms the total number of teachers that districts had hoped to hire*.
11. All estimates presented here were derived using the BRR weighting procedure utilized by the Wesvar 4.0 statistical analysis software as recommended by NCES. This method produced an estimate of 45,088 positions that were filled after the start of the school year, with a standard error of 529 and a

coefficient of variation of 1.173 percent.

12. Estimates are based on the public school district data set and do not include the responses from public charter schools.
13. In an ideal world, it would be possible to use the SASS data to estimate vacancies. In the school questionnaire, administrators were asked how difficult it was to fill vacancies across subject areas. Unfortunately, SASS did not ask *how many* positions schools were unable to fill, and it is not possible to estimate the difference between the number of teachers being sought at the start of the school year and the total hired after the year began. Since we are unable to calculate a vacancy rate, the analysis uses the late-fill rate as a proxy for the vacancy rate. Using this proxy assumes that there is a positive relationship between district late-fill rates and the likelihood that schools reported unfilled positions. This relationship was tested by regressing whether a school reported unfilled positions during the year (SASS School Respondents, question 36) against the district late-fill rate, and yielded a positive, significant correlation between the two variables. There is some evidence, therefore, supporting the use of the late-fill rate as a proxy.
14. The calculated estimates can be found in the appendix of this article.
15. The precise wording of the question was, "How difficult or easy was it to fill the vacancies for this school year in each of the following fields?" For different subject areas (e.g., General elementary, mathematics, special education, etc.), schools that had open positions could respond that it was "easy," "somewhat difficult" or "very difficult" to fill the position, or that the vacancy was never filled.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education.*

This article has been retrieved 1403 times since August 4, 2003

Volume 11 Number 24 **August 4, 2003** **ISSN 1068-2341**

High-Stakes Testing: Another Analysis

Barak Rosenshine
University of Illinois at Urbana, Champaign

Citation: Rosenshine, B. (2003, August 4). High-stakes testing: Another analysis. *Education Policy Analysis Archives, 11*(24). Retrieved [date] from http://epaa.asu.edu/epaa/v11n24/.

Abstract

Amrein and Berliner (2002b) compared National Assessment of Educational Progress (NAEP) results in high-stakes states against the national average for NAEP scores. They studied NAEP scores for 8th grade mathematics, 4th grade mathematics, and 4th grade reading. They concluded that states that introduced consequences (high-stakes) to their statewide tests did not show any particular gains in their statewide NAEP scores. However, there was no comparison group in their analysis. In this analysis, a comparison group was formed from states that did not attach consequences to their state-wide tests. This analysis showed that states that attached consequences outperformed the comparison group of states on each of the three NAEP tests for the last four-year period. These results showed that, overall, these was a meaningful carryover from attaching consequences on statewide tests to statewide NAEP scores.

As reported by Viadero (2003a, 2003b), Audrey Amrein and David Berliner (2002a, 2002b, 2002c) have studied the effects of attaching consequences or accountability ("high stakes") to student scores on their statewide exams. These consequences include monetary awards to schools or teachers, authority to replace a principal or a teacher, and limiting grade-to-grade promotion. Individual states imposed one to six such consequences, with an average of three consequences. (See Table 1 in Amrein and Berliner (2002b).)

In an admirable decision, Amrein and Berliner did not look at each state's scores on their own statewide tests. Rather, they looked at each state's scores on an independent measure, the National Assessment of Educational Progress. They compared the four-year changes in NAEP scores in each high-stakes state against the average changes for all the states that took each NAEP test. They studied NAEP results in three areas: 8th grade mathematics, 4th grade mathematics, and 4th grade reading.

They concluded that, there were "no consistent effects across states" after consequences were introduced. (Amrein & Berliner, 2002b, p. 57). Some states had larger increases than the national average, but the NAEP changes in other states were less than the national average. They concluded that students in high-stakes states were not learning anything beyond the specific content of the statewide tests.

Their analysis, however, did not include a comparison group. The economist John H. Bishop, quoted in Viadero (2003b), said "The natural thing to do would be to compare the states that had accountability systems to ones that didn't." This reanalysis follows Bishop's suggestion and is a comparison of the NAEP gains in the high-stakes states against the NAEP gains in states that did not have statewide accountability procedures.

Not all the 26 high-stakes states were included in the Amrein and Berliner analysis. They noted that some states may have manipulated their NAEP scores by exempting some of the special education students and students with limited English proficiency from those taking the NAEP test. They excluded these states or individual results from these states. For example, all the results from North Carolina and Texas, two states with large NAEP increases, were labeled as "unclear" and were not included in their analyses. All of their reported results are for the remaining 8 to 12 "clear" states (depending upon the specific NAEP exam). I believe this separation into "clear" and "unclear" high stakes states was a valuable step because it allows the analysis to focus only on the results in the clear states.

My analysis uses the NAEP four-year gain from these same 8-12 "clear" high-stakes states. An additional 14 to 18 states (depending upon the specific NAEP exam) did not attach consequences to tests and they served as the comparison group in my analysis. My analysis is based on the gain, from cohort to cohort, between 1996 and 2000 for the two mathematics tests and on the gain between 1994 and 1998 on the reading test. One could, of course, go back eight years, but then the number of high-stakes states would be much smaller. The tables from which these numbers came are readily available at the NAEP website (http://nces.ed.gov/nationsreportcard/).

The results in the Table 1 show the average NAEP increases in the "clear" high-stakes states were much higher than the increases in the comparison states. . In 8th grade mathematics and in 4th grade reading the mean increase for the clear high-stakes states was double the increase for the states without consequences. The effect sizes for the comparisons were .35 for 4th grade math, .79 for 8th grade math, and .61 for 4th grade reading. These effect sizes have been called moderate to large. An effect size of .35 means that the average high-stakes state would score at the 63rd percentile of the comparison states. Effect sizes of .79 and .62 correspond to the 78thand the 73rd percentile of the comparison states.

Table 1
Average four-year increases on NAEP scores for
Clear High-Stakes States and for States without High Stakes

NAEP Test	Average four-year increase in clear high-stakes states	Average four-year increase in states without high-stakes
4th grade mathematics (1996-2000)	3.45 (n=11 states)	2.40 (n=15 states)
8th grade mathematics (1996-2000)	3.42 (n=7 states)	1.63 (n=13 states)
4th grade reading (1994-1998)	3.44 (n=9 states)	1.21 (n=14 states)

These results suggest that students in the clear high-stakes states were, indeed, learning mathematics and reading that was beyond the specific content of the statewide tests. Their NAEP achievement, on all three tests, was higher than the achievement of students in NAEP states that did not attach high-stakes to their statewide exams.

These results might lead us to reconsider some of the textual statements in the Amrein and Berliner report (2002b). Amrein and Berliner wrote that "...the imposition of high-stakes testing results in a more narrow form of training...." (p. 6) Perhaps, and perhaps not. But this apparent narrow training did not prevent students in the clear high-stakes states from doing quite well on the NAEP tests, better than the students in states that did not attach consequences to statewide tests.

Did state NAEP scores really decrease?

The Amrein and Berliner (2002b) report states that NAEP scores in many high-stakes states "decreased" after consequences were implemented. They wrote that "grade 4 reading achievement decreased" (p.19) in Alabama after stakes were attached to statewide tests. They wrote that in Nevada, "grade 4 math achievement decreased." But "decrease" is a relative term. Scores in Nevada increased three points on the 4th grade NAEP math test between 1996 and 2000. But because this increase was less than the national average of four points, Nevada was listed as "grade 4 math achievement decreased." Grade 4 reading achievement increased by three points in Alabama between 1994 and 1998, the same increase as the national average. However, Amrein and Berliner noted that the percentage of students exempted from NAEP increased, in Alabama, between 1990 and 1994. Therefore, they concluded that "after stakes were attached to tests in Alabama, grade four reading achievement decreased." (p. 19)

Overall, Amrein and Berliner reported that NAEP scores in the 4th grade mathematics test "decreased" in 8 states. However, these was no actual decrease in any of these states: one state had no change between 1996 and 2000 but the remaining 7 states actually showed increases of from 1 to 4 points during this period. If one adds them up, across the three NAEP exams, there was an absolute decrease in only two of the high-stakes state scores used in the Amrein and Berliner analysis and an absolute decrease in 10 of the NAEP scores for states that did not include accountability measures.

Discussion

This analysis showed that the clear high-stakes states outperformed the comparison states on each of the three NAEP tests for the last four-year period. If my analytic approach makes sense, and if these results are confirmed by others, then I hope we can begin to study what these results mean

The results showed there was a meaningful carryover effect, in some states, from statewide testing to the NAEP. Based on these results, it is not appropriate for Amrein and Berliner (2002b) to say that attaching accountability to statewide tests "results in a narrow form of training," or "high-stakes testing creates a 'training effect' only." (p. 6) Nor is it appropriate to say that "students were learning the content of the state-administered test and perhaps little else." (p. 60)

Although attaching accountability to statewide tests worked well in some high-stakes states, it was not an effective policy in all states. South Carolina, Massachusetts, and Alabama did particularly well in 4th grade math, but not New Mexico, West Virginia, or Kentucky. Indiana and Alabama did particularly well in 8th grade math, but not New Mexico or Missouri. Louisiana, Delaware and Virginia did particularly well in 4th grade reading, but not Missouri or New Mexico.

It would be appropriate to study these successful and less successful high-stakes states and learn how they achieved their results. It would be less appropriate to simply use these results as a hammer and blindly require all states to impose consequences.

I find it unlikely that the NAEP results in the high-scoring states were obtained merely because two weeks were devoted to test preparation. I find it unlikely that the NAEP results in the high-scoring states were obtained only because consequences and accountability were introduced.

My guess would be that there is a strong academic focus in these classrooms and these schools. The research has supported academically-focused classrooms since 1960, and having seen lots of trivia in classrooms, I welcome a return to an academic focus. I've been in many low-income high-achievement elementary schools and they are, indeed, high achieving places. I have seen history projects, discussion of novels, Junior Great Books, impressive mathematics lessons, and tutoring during lunch hours. It would be unfair to the efforts of students, teachers, and principals to say that they are merely focusing on statewide tests.

Statewide policies and district policies may also facilitated these increases. If so, how was that accomplished? Did these policies affect all schools? What forms of help did the state and the district provide for the classrooms? All these items may be worthy of study. We might also reexamine the lists of statewide consequences in Table 1 of Amrein and Berliner (2002b) and ask which of these consequences might act as a motivating and not a threatening factor.

Audrey Amrein and David Berliner have performed an important service by focusing on the consequences that some states have attached to statewide testing. I think their use of state NAEP scores as an independent assessment was a brilliant move. Their use of NAEP scores also allows others to conduct additional analyses of this public data. My additional analysis suggests that students in some high-stakes states have done very well on the NAEP tests. I hope we can now study how this happened.

Acknowledgement

I was in contact with Audrey Amrein throughout this reanalysis and it could not have been done without her help. She provided clear and full answers to each of my frequent email questions. . Some of my ideas for this analysis came as a result of email discussions with David Berliner, and I thank him. Sam Stringfield, Burnie Bond, Bob Stevens, Jere Brophy and Marilyn Kohl all provided useful questions and comments.

References

Amrein, A.L. & Berliner, D.C. (2002a, March 28). High-stakes testing, uncertainty, and student learning *Education Policy Analysis Archives*, *10*(18). Retrieved July 18, 2003 from http://epaa.asu.edu/epaa/v10n18/.

Amrein, A.L. & Berliner, D.C. (2002b). The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP Test results in states with high school graduation exams . Tempe, AZ: Education Policy Studies Laboratory, Arizona State University. Retrieved July 18, 2003 from http://www.asu.edu/educ/epsl/EPRU/documents/EPSL-0211-126-EPRU.pdf.

Amrein, A.L. & Berliner, D.C. (2002c). An analysis of some unintended and negative consequences of high-stakes testing. Tempe, AZ: Education Policy Studies Laboratory, Arizona State University. Retrieved July 18, 2003 from http://www.asu.edu/educ/epsl/EPRU/documents/EPSL-0211-125-EPRU.pdf.

Viadero, D. (2003a, January 8). Reports find fault with high-stakes testing. *Education Week, 22*(16), 5.

Viadero, D. (2003b, February 5). Researchers debate impact of tests. *Education Week, 22*(21), 1 & 12.

About the Author

Barak Rosenshine
College of Education
University of Illinois, Urbana—Champaign

Email: rosenshine@uiuc.edu

Barak Rosenshine is an emeritus professor of educational psychology at the University of Illinois at Urbana-Champaign. His research speciality is classroom instruction.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 802 times since August 4, 2003

Volume 11 Number 25 **August 4, 2003** **ISSN 1068-2341**

Re-analysis of NAEP Math and Reading Scores in States with and without High-stakes Tests: Response to Rosenshine

Audrey Amrein-Beardsley
David C. Berliner
Arizona State University

Citation: Amrein-Beardsley, A. A. & Berliner, D. C. (2003, August 4). Re-analysis of NAEP math and reading scores in states with and without high-stakes tests: Response to Rosenshine. *Education Policy Analysis Archives, 11*(25). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n25/.

Abstract

Here we address the criticism of our NAEP analyses by Rosenshine (2003). On the basis of his thoughtful critique we redid some of the analyses on which he focused. Our findings contradict his. This is no fault of his, the reasons for which are explained in this paper. Our findings do support our position that high-stakes tests do not do much to improve academic achievement. The extent to which states with high-stakes tests outperform states without high-stakes tests is, at best, indeterminable. Using 1994-1998 NAEP reading and 1996-2000 NAEP math data and accounting for NAEP exemption rates for the same years, we found that states with high-stakes tests are not outperforming states without high-stakes tests in reading in the 4th grade or math in the 8th grade at a statistically significant level. States with high-stakes tests are, however, outperforming states without high-stakes tests in math in the 4th grade at a statistically significant level. Our findings also support our earlier stance that states with high-stakes tests are exempting more students from participating in the NAEP than are states without high-stakes tests. This is more prevalent the more recent the NAEP test administration. This is illustrated in the tables below.

Introduction

In our research, we were concerned that scores on high-stakes state tests could easily be manipulated through narrowing of the curriculum, drilling on items similar to the test, increasing exclusion rates of students, increasing in dropouts and push-outs, and the like. To judge whether that concern was valid, we looked at audit tests—tests that might have some overlap with a state's own test but where school personnel were under much less intense pressure to achieve higher scores. We chose a series of audit tests to examine—SAT, ACT and AP tests, as well as all administrations of the NAEP reading and mathematics tests. We also studied whether some unanticipated side effects were present when high-stakes tests were introduced, such as increased GED taking, increased reporting of cheating, problems of teacher morale, problems with student motivation to learn, and so forth.

Substantive criticism of our work, thus far, has been limited to the NAEP data we reported. To our knowledge, the other conclusions we reached have not yet been subject to the same kinds of thoughtful criticism. So for now, given the methods that we used for analyses, our findings in those other areas stand. We concluded that there was no systematic pattern of gains on SATs, ACTs or AP exams. That is, we found no evidence of transfer from the state tests to these other tests, tests that can be considered as audit measures. In addition, we found increased drop-out rates and decreased high school graduation rates, increased rates by which students participated in the GED program, and a host of troubling negative affects associated with high-stakes testing.

Here we address the criticism of our NAEP analyses by Rosenshine (2003). On the basis of his thoughtful criticism, we redid some of the analyses on which he focused and now have a different view of the findings. What we found contradicts what we found in both of our earlier papers (Amrein & Berliner, 2002a; Amrein & Berliner 2002b) but the data analyzed below are for different years of data from those used in the earlier papers. Following the form of the analyses done by Rosenshine the data analyzed below are only for the years 1994-1998 for the NAEP reading test, and 1996-2000 for the NAEP mathematics tests. In addition, in our earlier work we used the national trend line as the contrast or control group for our analyses. In this analysis, we use the composite score for states without high-stakes tests as the control. In addition, our findings contradict the findings reported by Rosenshine. This is no fault of his. Rosenshine used our designation of clear and unclear states with and without high-stakes tests from the second of our two papers.[2] We communicated many times and approved the states he used in his analysis. Given more consideration, however, we noticed the distinctions we made between clear and unclear states was based on our overall findings which were based on all of the available NAEP data. In other words, Rosenshine analyzed the latest two NAEP administrations in reading and math using the distinctions we made between clear and unclear states when we used all of the available NAEP data, approximately 10 years of NAEP data per subject. To complicate things more, because we used the national trend line as our control group, our clear/unclear distinctions were also made factoring in the national average. Rosenshine did not do this which makes for differences in the findings. He used the states without high-stakes tests as the control. This makes for a better analysis and we have followed his lead here. In short, Rosenshine should not be faulted for his findings nor should he be considered wrong in what he did. He did a fine reanalysis of our NAEP examination given the information he had at that point, and here we are redoing his.

NAEP Reading Grade 4 1994-1998

Taking Table 1 from Amrein & Berliner (2002b) and the states in which high stakes tests were implemented before 1994 and between 1994 and 1998, we re-ran our analyses, as Rosenshine did, using all states with high-stakes tests and, as the control group, all states without high-stakes tests for which NAEP data were available. What we found in regards to reading grade 4 achievement from 1994-1998 is as follows:

Table 1

Fourth grade 1994-1998 NAEP reading scores (raw data).

States without high-stakes tests:	NAEP 1994	NAEP 1998
Arizona	206	207
Arkansas	209	209
California	197	202
Colorado	213	222
Connecticut	222	232
Delaware	206	212
Florida	205	207
Georgia	207	210
Hawaii	201	200
Iowa	223	223
Kansas	n/a	222
Maine	228	225
Massachusetts	223	225
Minnesota	218	222

States with high-stakes tests:	NAEP 1994	NAEP 1998
Alabama	208	211
Kentucky	212	218
Louisiana	197	204
Maryland	210	215
Michigan	n/a	217
Mississippi	202	204
Missouri	217	216
New Mexico	205	206
North Carolina	214	217
Oklahoma	n/a	220
South Carolina	203	210
Tennessee	213	212
Texas	212	217
West Virginia	213	216

Montana	222	226					
Nevada	n/a	208					
New Hampshire	223	226					
New York	212	216					
Oregon	n/a	214					
Rhode Island	220	218					
Utah	217	215					
Virginia	213	218					
Washington	213	217					
Wisconsin	224	224	Change in Score				Change in Score
Wyoming	221	219					
OVERALL	214.7	216.8	+2.1*	OVERALL	208.8	213.1	+4.3*

*Significant at a p < .05 level

Table 1 illustrates that the states with high-stakes tests outperformed those states without high-stakes tests on the NAEP grade 4 reading tests over the period 1994-1998. However, as shown in our earlier research (Amrein & Berliner, 2002a; Amrein & Berliner, 2002b), the rates by which students are excluded from the NAEP must be taken into consideration to determine whether gains and losses are clear (interpretable) or unclear (not interpretable).

Clear gains can be determined if a state's scores increase while the rates by which students are exempted from the NAEP stay the same or decrease. In other words, when the pool of students sampled to participate in the NAEP is less selective then the likelihood that their scores would increase artificially is nullified. Under these conditions such gains are clear. Clear losses can be determined if a state's scores decrease at the same time the rates by which students are exempted from the NAEP increase. In this case, the pool of students sampled was more selective and yet the scores still went down. Under these conditions it is reasonable to interpret those findings as a clear loss.

Unclear gains are the case when a state's scores increase while the rates by which students are exempted from the NAEP increase. In other words, the pool of students sampled to participate in the NAEP is more selective and therefore likely to have biased the resulting gains. If lower-scoring students are pulled from the NAEP sample, scores on the NAEP will increase. This makes for unclear results. Unclear losses are the case when a state's scores decrease at the same time the rates by which students are exempted from the NAEP sample decrease. In this case, the pool of students sampled was less selective so it is difficult to determine whether the addition of more lower-scoring students or an actual decrease in achievement caused the resulting losses.

We believe that it is absolutely necessary to make these kinds of judgments about each state because states with high-stakes tests are those states that increasingly are exempting more students from participating in the NAEP. "In states with high-stakes tests, between 0%–49% of the gains in NAEP scores can be explained by increases in rates of exclusion." (Amrein & Berliner, 2002a)

Looking simply at those states for which clear gains or losses are applicable, an analysis of the data yields the results given in Table 2. In this table states shaded in green are those for which clear results were evident, states shaded in red are those for which unclear results were illustrated, and states shaded in yellow are those for which there were not enough data to analyze gains or losses appropriately.

As can be seen, only two states included in the states with high-stakes column can be counted as states with "clear" effects. The composite data are not significant but the table illustrates the extent to which states with high-stakes tests are not gaining in score simply because of their high-stakes testing policies.

Table 2

Fourth grade 1994-1998 NAEP reading scores with states coded as clear or unclear in their gains and losses.

States without high-stakes tests:	NAEP 1994	NAEP 1998		States with high-stakes tests:	NAEP 1994	NAEP 1998	
Arizona	206	207		Alabama	208	211	
Arkansas	209	209		Kentucky	212	218	
California	197	202		Louisiana	197	204	
Colorado	213	222		Maryland	210	215	
Connecticut	222	232		Michigan	n/a	217	
Delaware	206	212		Mississippi	202	204	
Florida	205	207		Missouri	217	216	
Georgia	207	210		New Mexico	205	206	
Hawaii	201	200		North Carolina	214	217	
Iowa	223	223		Oklahoma	n/a	220	
Kansas	n/a	222		South Carolina	203	210	
Maine	228	225		Tennessee	213	212	
Massachusetts	223	225		Texas	212	217	
Minnesota	218	222		West Virginia	213	216	
Montana	222	226					
Nevada	n/a	208					
New Hampshire	223	226					
New York	212	216					
Oregon	n/a	214					
Rhode Island	220	218					
Utah	217	215					
Virginia	213	218					
Washington	213	217					
Wisconsin	224	224	Change in Score				Change in Score
Wyoming	221	219					
OVERALL	215.4	217.0	+1.6*	OVERALL	209.5	210.0	+0.5

*Significant at a p < .05 level

The composite data are important in that they nullify what one might conclude looking simply at Table 1. States with high-stakes tests are not outperforming states without high-stakes tests in reading grade 4 performance. Rather, as illustrated in Table 2, states without high-stakes tests gained in reading grade 4 performance at a statistically significant level. Given only two states are included as clear states with high-stakes tests, states with high-stakes tests made insignificant gains and the differences between the two mean gains are not statistically significant.

Most importantly, what can be drawn from Table 2 is that states with high-stakes tests are exempting more students from participating in the reading grade 4 NAEP. Ninety percent of the states with "unclear" gains are states with increases in the rates by which students were exempted from the test. This supports the notion that states with high-stakes tests are not gaining in NAEP scores simply because of their high-stakes testing policies.

NAEP Math Grade 4 1996-2000

Taking Table 1 from Amrein & Berliner (2002b) and the states in which high stakes tests were implemented before 1996 and between 1996 and 2000, we re-ran our analyses using all states with high-stakes tests and, as the control group, all states without high-stakes tests for which NAEP data were available. What we found in regards to math grade 4 achievement from 1996-2000 is as follows:

Table 3

Fourth grade 1996-2000 NAEP mathematics scores (raw data)

States without high-stakes tests:	NAEP 1996	NAEP 2000		States with high-stakes tests:	NAEP 1996	NAEP 2000	
Alaska	224	n/a		Alabama	212	218	
Arizona	218	219		California	209	214	
Arkansas	216	217		Delaware	215	n/a	
Colorado	226	n/a		Florida	216	n/a	
Connecticut	232	234		Indiana	229	234	
Georgia	216	220		Kentucky	220	221	
Hawaii	215	216		Louisiana	209	218	
Idaho	n/a	227		Maryland	220	222	
Illinois	n/a	225		Massachusetts	229	235	
Iowa	229	233		Michigan	226	231	
Kansas	n/a	232		Mississippi	208	211	
Maine	233	231		Missouri	225	229	
Minnesota	232	235		Nevada	217	220	
Montana	228	230		New Jersey	227	n/a	
Nebraska	228	226		New Mexico	214	214	
North Dakota	231	231		New York	223	227	
Oregon	224	227		North Carolina	224	232	
Rhode Island	221	225		Ohio	n/a	231	
Tennessee	219	220		Oklahoma	n/a	225	
Utah	226	227		Pennsylvania	226	n/a	
Vermont	225	232		South Carolina	213	220	
Washington	225	n/a		Texas	229	233	
Wisconsin	231	n/a	Change in Score	Virginia	222	230	Change in Score
Wyoming	223	229		West Virginia	224	225	
OVERALL	224.9	226.8	+1.9*	OVERALL	219.9	224.5	+4.6*

*Significant at a $p < .05$ level

Table 3 illustrates that the states with high-stakes tests outperformed those states without high-stakes tests on the math NAEP grade 4 tests over the time period 1996-2000. However, as argued earlier, the rates by which students are excluded from the NAEP must be taken into consideration to determine whether gains and losses are clear or unclear. Using the same rules as outlined above to determine clear and unclear gains and losses, we looked at only those states for which clear gains or losses are relevant. (For 4th grade reading, Rosenshine included the following states: Arizona, Arkansas, California, Connecticut, Hawaii, Iowa, Maine, Montana, New Hampshire, Rhode Island, Utah, Washington, Wisconsin, and Wyoming. There are notable differences in the states he included and the states we included that likely came from the fact that we drew our states directly out of Table 1 of the original document.) An analysis of the data yields the following:

Table 4

Fourth grade 1996-2000 NAEP mathematics scores with states coded as clear or unclear in their gains and losses.

States without high-stakes tests:	NAEP 1996	NAEP 2000		States with high-stakes tests:	NAEP 1996	NAEP 2000	
Alaska	224	n/a		Alabama	212	218	
Arizona	218	219		California	209	214	
Arkansas	216	217		Delaware	215	n/a	
Colorado	226	n/a		Florida	216	n/a	
Connecticut	232	234		Indiana	229	234	
Georgia	216	220		Kentucky	220	221	
Hawaii	215	216		Louisiana	209	218	
Idaho	n/a	227		Maryland	220	222	
Illinois	n/a	225		Massachusetts	229	235	
Iowa	229	233		Michigan	226	231	
Kansas	n/a	232		Mississippi	208	211	
Maine	233	231		Missouri	225	229	
Minnesota	232	235		Nevada	217	220	
Montana	228	230		New Jersey	227	n/a	
Nebraska	228	226		New Mexico	214	214	
North Dakota	231	231		New York	223	227	
Oregon	224	227		North Carolina	224	232	
Rhode Island	221	225		Ohio	n/a	231	
Tennessee	219	220		Oklahoma	n/a	225	
Utah	226	227		Pennsylvania	226	n/a	
Vermont	225	232		South Carolina	213	220	
Washington	225	n/a		Texas	229	233	
Wisconsin	231	n/a	Change in Score	Virginia	222	230	Change in Score
Wyoming	223	229		West Virginia	224	225	
OVERALL	224.5	225.6	+1.1	OVERALL	210.4	215.0	+4.6*

*Significant at a p < .05 level

Compared to the reading data above, we now find the opposite when we look at the math grade 4 NAEP composite data. When states with clear effects are pulled out and analyzed, it is apparent that states with high-stakes tests are outperforming states without high-stakes tests at a statistically significant level. The scores posted by the clear states with high-stakes tests are significantly different than the scores posted by the clear states without high-stakes tests.

Again, however, what can also be drawn from Table 4 is that states with high-stakes tests are exempting more students from participating in the math grade 4 NAEP. Two times as many states with high-stakes tests exempted students and realized gains in grade 4 math achievement from 1996-2000 than did states without high-stakes tests. This, again, supports the notion that states with high-stakes tests are not all gaining in NAEP scores simply because of their high-stakes testing policies.

NAEP Math Grade 8 1996-2000

Taking Table 1 from Amrein & Berliner (2002b) and the states in which high stakes tests were implemented before 1996 and between 1996 and 2000, we re-ran our analyses using all states with high-stakes tests and, as the control group, all states without high-stakes tests for which NAEP data were available. What we found in regards to math grade 8 achievement from 1996-2000 is as follows:

Table 5

Eighth grade 1996-2000 NAEP mathematics scores (raw data)

States without high-stakes tests:	NAEP 1996	NAEP 2000		States with high-stakes tests:	NAEP 1996	NAEP 2000	
Alaska	278	n/a		Alabama	256	262	
Arizona	268	271		California	263	262	
Arkansas	261	261		Delaware	267	n/a	
Colorado	276	n/a		Florida	264	n/a	
Connecticut	280	282		Indiana	275	283	
Georgia	262	266		Kentucky	267	272	
Hawaii	262	263		Louisiana	252	259	
Idaho	n/a	278		Maryland	270	276	
Illinois	n/a	277		Massachusetts	277	283	
Iowa	284	n/a		Michigan	276	278	
Kansas	n/a	284		Mississippi	250	254	
Maine	284	284		Missouri	274	274	
Minnesota	284	288		Nevada	n/a	268	
Montana	283	287		New Mexico	262	260	
Nebraska	283	281		New York	270	276	
North Dakota	284	283		North Carolina	268	280	
Oregon	277	281		Ohio	n/a	283	
Rhode Island	268	273		Oklahoma	n/a	272	
Tennessee	263	263		South Carolina	260	266	
Utah	276	275		Texas	270	275	
Vermont	279	283		Virginia	270	277	
Washington	276	n/a		West Virginia	265	271	
Wisconsin	283	n/a	Change in Score				Change in Score
Wyoming	275	277					
OVERALL	275.5	276.7	+1.2*	OVERALL	266.1	271.6	+5.4*

*Significant at a p < .05 level

Table 5 illustrates the states with high-stakes tests outperformed those states without high-stakes tests on the math NAEP grade 8 1996-2000. Again, we argue that the rates by which students are excluded from the NAEP must be taken into consideration to determine whether gains and losses are clear or unclear.

Using the same rules as outlined above to determine clear and unclear gains and losses, we looked at only those states for which clear gains or losses are apparent (Note 3). An analysis of the data yields the following:

Table 6

Eighth grade 1996-2000 NAEP mathematics scores with states coded as clear or unclear in their gains and losses.

States without high-stakes tests	NAEP 1996	NAEP 2000	Change in Score	States with high-stakes tests	NAEP 1996	NAEP 2000	Change in Score
Alaska	278	n/a		Alabama	256	262	
Arizona	268	271		California	263	262	
Arkansas	261	261		Delaware	267	n/a	
Colorado	276	n/a		Florida	264	n/a	
Connecticut	280	282		Indiana	275	283	
Georgia	262	266		Kentucky	267	272	
Hawaii	262	263		Louisiana	252	259	
Idaho	n/a	278		Maryland	270	276	
Illinois	n/a	277		Massachusetts	277	283	
Iowa	284	n/a		Michigan	276	278	
Kansas	n/a	284		Mississippi	250	254	
Maine	284	284		Missouri	274	274	
Minnesota	284	288		Nevada	n/a	268	
Montana	283	287		New Mexico	262	260	
Nebraska	283	281		New York	270	276	
North Dakota	284	283		North Carolina	268	280	
Oregon	277	281		Ohio	n/a	283	
Rhode Island	268	273		Oklahoma	n/a	272	
Tennessee	263	263		South Carolina	260	266	
Utah	276	275		Texas	270	275	
Vermont	279	283		Virginia	270	277	
Washington	276	n/a		West Virginia	265	271	
Wisconsin	283	n/a					
Wyoming	275	277					
OVERALL	271.1	271.9	+0.7	OVERALL	258.8	261.8	+3.0

After the states with clear effects are pulled out and analyzed, it seems that states with high-stakes tests are outperforming states without high-stakes tests. They are not, however, outperforming states without high-stakes tests at a statistically significant level. In addition, the scores posted by the clear states with high-stakes tests are not significantly different than the scores posted by the clear states without high-stakes tests. States with high-stakes tests are not outperforming states without high-stakes tests in math grade 8 performance.

Again, what can also be drawn from Table 6 is that states with high-stakes tests are exempting more students from participating in the math grade 8 NAEP. Thirty-three percent of the states without high-stakes tests exempted more students and realized gains in math grade 8 NAEP scores. Fifty percent of the states with high-stakes tests exempted more students and realized gains in math grade 8 NAEP scores. This, again, supports our assertion that states with high-stakes tests are not gaining in NAEP scores simply because of their high-stakes testing policies.

Conclusion

In short, states with high-stakes tests seem to have outperformed states without high-stakes tests on the grade 4 math NAEP at a statistically significant level. However, gains between states with and without high stakes tests were not statistically different on the grade 4 reading or the grade 8 math NAEP. States with high-stakes tests are *not* outperforming states without high-stakes tests on both of these measures.

In addition, the rates by which personnel in states with high-stakes tests are exempting students are increasing at a faster rate than they are in states without high-stakes tests. There may be an underlying characteristic other than high-stakes tests that is causing this phenomenon, but this would take further analyses. What we do know, however, is that for the most part the gains posted by states with high-stakes tests on two of the three NAEP tests are more related to the rates by which students are exempted from the tests than they are related to high-stakes tests themselves.

We thank Professor Rosenshine for suggesting these alternative analytic techniques to us. In the end, for now, we remain unconvinced that the NAEP tests are showing much in the way of transfer effects. Given all the data we reported in our previous reports we remain unconvinced that the high-stakes tests used by states are showing systematic positive affects on audit tests used to assess transfer.

References

Amrein, A.L. & Berliner, D.C. (2002a, March 28). High-stakes testing, uncertainty, and student learning *Education Policy Analysis Archives*, *10*(18). Retrieved July 24, 2003 from http://epaa.asu.edu/epaa/v10n18/.

Amrein, A.L. & Berliner, D.C. (2002b). The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP Test results in states with high school graduation exams. Tempe, AZ: Education Policy Studies Laboratory, Arizona State University. Retrieved July 24, 2003 from http://www.asu.edu/educ/epsl/EPRU/documents/EPSL-0211-126-EPRU.pdf.

Rosenshine, B. (2003, August 4). High-stakes testing: Another analysis. *Education Policy Analysis Archives*, *11* (24). Retrieved August 4, 2003 from http://epaa.asu.edu/epaa/v11n24/.

About the Authors

Audrey Amrein-Beardsley
Email: audrey.beardsley@cox.net

Audrey Amrein-Beardsley is a part-time Research and Evaluation Associate at a private foundation in Scottsdale, Arizona and a part-time researcher at Arizona State University. She received her PhD from Arizona State University in 2002 in Education Policy with an emphasis in Research Methodology. Her scholarly interests include the study of the intended and unintended consequences of high-stakes testing policies. To date, she has focused on the effects of high-stakes testing policies at a macro level given the frequency with which these policies are being implemented across the nation.

David C. Berliner
Regents' Professor of Education
College of Education
Arizona State University
Tempe, AZ 85287-2411

Email: berliner@asu.edu

David C. Berliner is Regents' Professor of Education at the College of Education of Arizona State University, in Tempe, AZ. He received his Ph.D. in 1968 from Stanford University in educational psychology, and has worked also at the University of Massachusetts, WestEd, and the University of Arizona. He has served as president of the American Educational Research Association (AERA), president of the Division of Educational Psychology of the American Psychological Association, and as a fellow of the Center for Advanced Study in the Behavioral Sciences and a member of the National Academy of Education. Berliner's publications include *The Manufactured Crisis*, Addison-Wesley, 1995 (with B.J. Biddle) and *The Handbook of Educational Psychology*, Macmillan, 1996 (Edited with R.C. Calfee). Special awards include the Research into Practice Award of AERA, the National Association of Secondary School Principals Distinguished Service Award, and the Medal of Honor from the University of Helsinki. His scholarly interests include research on teaching and education policy analysis.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona

casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil
(AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 1019 times since August 6, 2003

Volume 11 Number 26 **August 6, 2003** **ISSN 1068-2341**

The Role of Theory and Policy in the Educational Treatment of Language Minority Students: Competitive Structures in California

Tom T. Stritikus
University of Washington

Eugene Garcia
Arizona State University

Citation: Stritikus, T. T. & Garcia, E. (2003, August 6). The role of theory and policy in the educational treatment of language minority students: Competitive structures in California. *Education Policy Analysis Archives, 11*(26). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n26/.

Abstract

For teachers, theories play a central role in guiding the intellectual work that they have chosen to perform. Teachers are guided by both theories which they use to interpret, analyze and take action in their professional worlds. At any given time, teachers may be faced with multiple and competing theoretical perspectives which attempt to influence their classroom practice. In this article, we examine the theoretical and policy-based positions currently competing to shape the nature of educational practice for language minority students. We highlight the salient theoretical differences between *additive* and *subtractive conceptions* for the education of language minority students and their policy- and practice-based implications. Then, we examine select findings from one district's implementation of Proposition 227, and consider how teachers react when competing theories attempt to shape their classroom practice. Specifically, we consider: How might teachers' theories be complemented or contrasted by the underlying theoretical position of Proposition 227? How do teachers' theories about their students mediate the manner in which they react and respond to the policy shift away from native language instruction? We conclude by considering what implications *additive* and *subtractive* competitive structures have for the future of policy and practice for language minority students in the United States.

In their practice, teachers are like other learners in the sense that they interpret new ideas and attempts to change their practice based on their existing understandings. The manner in which teachers modify new ideas is based upon their guiding extant theories about their profession and their students (Cohen & Barnes, 1993; Kennedy, 1991; Woods, 1994).

The importance of teachers' theories and world views has been highlighted by empirical research. In a recent study of effective teachers for language minority students, teachers reported that they have very well articulated theories of how children develop and learn and the role education plays in such processes (Garcia, 1999). In short, whether we articulate them or not, we all have theories that guide us in making meaning of the world we live in.

Consequently, it is important to look more explicitly at conceptual frameworks—theories—which might help us understand the educational circumstances of language minority students and their teachers. This is particularly the case when a host of competing theories attempt to drive policy and practice for language minority students. Such is the case with California's Proposition 227 and its attempt to end bilingual education in that state. Specifically, we consider the following two questions: First, how might teachers' theories be complemented or contrasted by the underlying theoretical position of Proposition 227? Second, how do teachers' theories about their students mediate the manner in which they react and respond to policy shift away from native language instruction? To consider these questions, we examine the theoretical and policy-based positions currently competing to shape the nature of educational practice for language minority students. Then, to consider the empirical implications of these questions, we examine select findings from Walton Unified School District's implementation of Proposition 227. We use the experiences of the three teachers' from the small rural district to illustrate how teachers' theories regarding the needs of their students, bilingual education, and language maintenance influenced their reaction to Proposition 227. (For a full discussion of Proposition 227 implementation in Walton Unified see Stritikus (2002).) We conclude by considering what implications competitive structures have for the future of policy and practice for language minority students in the United States.

Competing Theories for the Education Language Minority Students

Proposition 227, known by its proponents as the "English for the Children Initiative," passed with a 61% majority of California voters on June 2, 1998. The initiative was an example of "people making law," written in response to apparent widespread discontent with the state's theories/policies regarding the education of non-English speaking children in public schools. Its intent was to inject more English instruction for these students in California's public schools. Some 25% of California's students currently fall into this student category and are referred to as Limited English Proficient (LEP), English Language Learners (ELL), and/or as language minority students. The assumption which lay under the initiative was that teaching children in their native language served only to hold them back in their acquisition of English and therefore in their future educational success.

Immediately upon its passage, Proposition 227 became a part of the California Education Code (#300-340). As it required within its text, districts throughout the state were given only 60 days to implement it. Under this new education code, children entering California Public Schools with very little English must be "observed" for a period of 30 calendar days. After 30 days, school personnel must decide if children have enough fluency in English to manage in a mainstream English classroom. If not, they are eligible to receive one year of "Sheltered English Immersion," also referred to as "Structured English Immersion," a program of English language instruction not described in detail in the law except to require that instruction be "nearly all" in English (with a definition for the term "nearly all" left up to the district's discretion). After one year, children are normally expected to integrate into mainstream English classrooms, where instruction is required to be "overwhelmingly" in English (again, with a definition for the term "overwhelmingly" left up to the district's discretion). If parents or legal guardians find that district or school personnel, including classroom teachers, "willfully and repeatedly refuse" to provide the English instruction as required, they have the right to sue for damages. This aspect of the law has not yet been fully tested in the courts.

The only legal alternative to placing an ELL student in a Sheltered English Immersion and/or mainstream English classrooms is the utilization of the parental waiver process. According to the new law, children who have special language needs, or whose parents specifically request it, can be placed in "Alternative Programs," most likely some form of bilingual program which includes instruction in the child's primary language. In order for a child to be enrolled in such a program, the parent or guardian must visit the school annually and sign a waiver requesting the placement. However, the first year a child enters California schools s/he must go through 30 days of "observation," generally conducted in English language classrooms, even if s/he has a signed waiver. Once the 30 days is completed, the child can enroll in an alternative program.

Along with the changes outlined above, the law allocates $50,000,000 per year to train adult English learners, parents or members of the community, to serve as tutors for children learning English. Finally, the new law is careful to state that if any conflicts are uncovered between its requirements and federal, state or constitutional law, those conflicts are resolved by following the "higher authority" of that previous law.

There are some areas of the California State Board of Education's policy regarding the instruction of Language Minority children that were not at all affected by the passage of Proposition 227. Teacher credentialing has remained the same, as have the requirements regarding the assessment of LEP children in English and in their native language. It is still required by law that schools and districts communicate with language minority families in their primary language whenever necessary. Children who are identified as in need of Special Education and operate under an Individual Education Plan are not touched by the changes.

Proposition 227 and the Move to Toward Subtractive Schooling

Proposition 227 certainly altered basic elements of policy toward language minority children in California's public schools. There had been a twenty-year tradition, thorough legislative and executive actions, encouraging, even mandating bilingual education programs in California. In 1987, these laws officially sunset, leaving districts less clear on the mandate from the state. Nonetheless, even since 1987, there had been a climate of increasing openness toward bilingual programs and other special services for language minority students among California school districts. Although state level support of bilingual education existed, multiple districts and schools across the state had taken steps to limit or weaken bilingual education programs (Wong Fillmore, 1992).

Bilingual education is not, and never has been, a neutral process. The education of linguistically diverse students is situated in larger issues about immigration, distribution of wealth and power, and the empowerment of students (Cummins, 2000; Heller, 1994). Policy and practice questions are situated in debates surrounding the legitimacy of the language and culture of diverse groups (Olsen, 1997). The subtractive and additive frameworks advanced in this literature review offer a way to situate the nature of teacher theories, educational practice, and educational policy in these broader debates surrounding the place of culturally and linguistically diverse students in the United States. A debate intensified by the changes in the California Education Code brought about by the voter initiative Proposition 227 and its reversal of the state's official support of primary language instruction.

Garcia (1995) and Garcia and Gonzalez (1995) serve as exemplars of the theoretical/policy/practice position that was overturned by Proposition 227. Imbedded in this additive perspective for language minority students is the understanding that language, culture, and their accompanying values, are constructed in the home and community environments, that children come to school with some constructed knowledge about many things, and that children's development and learning is best understood as the interaction of previous and present linguistic, socio-cultural, and cognitive constructions. An appropriate perspective of teaching language minority students is one that recognizes that learning becomes enhanced when it occurs in contexts that are socio-culturally, linguistically and cognitively meaningful for the learner (Garcia, 1995; Moll, 1994). Moreover, policies should reflect these conceptual underpinnings. It was the case that re-authorization of federal policy did exactly that, recognizing the importance of native language instruction and supporting those programs that were additive in nature (Garcia and Gonzalez, 1995; Wiese and Garcia, 1998). Table 1 exemplifies the attributes of school-wide and teacher practices associated with this conceptual framework. This is clearly contrasted with the conceptual framework that is at the foundation of Proposition 227: a disregard for non-English skills and circumstances outside of school and a focus on the instruction of English in English. Table 2 articulates the school-wide practices and teacher practices following from this conceptual framework. The distinction between additive and subtractive conceptions of cultural and linguistic diversity is not meant to be a strict dichotomy of policies and practices, but rather a framework for understanding the range of possible educational alternatives which exist for cultural and linguistically diverse students.

Table 1. Additive Conceptual Dimensions of Addressing Cultural and Linguistic Diversity

School-wide Practices

- A vision defined by the acceptance and valuing of diversity--Americanization is NOT the goal
- Professional development characterized by collaboration, flexibility and continuity with a focus on teaching, learning and student achievement
- Elimination (gradual or immediate) of policies that seek to categorize diverse students thereby rendering their educational experiences as inferior or limiting for further academic learning
- Reflection of and connection to surrounding community--particularly with the families of the students attending the school

---

Teacher Practices

- Bilingual/bicultural skills and awareness
- High expectations of diverse students
- Treatment of diversity as an asset to the classroom
- Ongoing professional development on issues of cultural and linguistic diversity and practices that are most effective
- Basis of curriculum development to address cultural and linguistic diversity
- Attention to and integration of home culture/practices
- Focus on maximizing student interactions across categories of Spanish and English proficiency and academic performance
- Focus on language development through meaningful interactions and communications

Table 2. Subtractive Conceptual Dimensions of Addressing Cultural and Linguistic Diversity

School-wide Practices

- A vision defined by the learning of English--Americanization/assimilation is the goal
- Professional development characterized a focus on direct teaching, emphasizing instruction of phonology, grammar and phonics in reading
- Elimination (gradual or immediate) of policies that seek to provide special instruction to a category of students marked by their non-English proficiency
- Connection to surrounding community--particularly with the families of the students attending the school emphasizing the development and use of English

---

Teacher Practices

- English development skills and awareness
- Expectations that English proficiency by students will enhance academic achievement
- Treatment of linguistic diversity as a characteristic that must be minimized
- Ongoing professional development and direct enforcement of direct teaching practices
- Basis of curriculum development to address cultural and linguistic assimilation
- Attention to and integration of diverse cultures into the "norm"
- Focus on maximizing student academic English development as assessed by English language development and academic testing--in many cases, "high stakes" testing
- Focus on English language, reading and literacy development through methods of direct instructions of skills

The subtractive position advanced by Proposition 227—as summarized by the practices embodied in Table 2—is contrasted by multiculturalist and multilingualist notions that English-only instruction is deeply problematic. Rather than view the home language and culture through a lens of deficit, multiculturalist and additive perspectives urge schools to see these as valuable educational resources. (Banks, 1995; Garcia, 1999; Gutiérrez, et al., 2000; Olneck, 1995). Proposition 227 presents a direct challenge to the notion that languages other than English have a legitimate and valuable place in the education of diverse students. Hence, the normative assumptions underlying Proposition 227 position the language and culture of diverse students in a subordinate and inferior role to English (Auerbach, 1995; Cummins, 2000; Kerper-Mora, 2000).

These normative assumptions have important consequences that extend beyond the classroom. The nature of the law works to position certain groups in a peripheral role in American society. Sekhon (1999), in an article assessing the legal and political implications of the proposition, argues that Proposition 227 positions immigrants on the outside of mainstream America:

> Proposition 227 positions English as "our" language by constructing it as our unlearned capacity: It is our birthright. The proposition differentiates "us" from "them" by denominating them in terms of an essential inability to call English their own. They must learn it. Proposition 227 not only demands that they learn our language, it demands that they forget their own. In so demanding, the proposition not only unleashes a salvo in the bilingual education debate, but is a crucial moment in the broader debate over assimilation and acculturation. (p. 1445)

Thus, in its scope, focus, and ideological implications, Proposition 227 differs markedly from past educational reforms. Teachers were not only told to shift educational practice, but forced to participate in an evolving debate about which theory would hold prominence in the education of language minority students. The distinction between additive and subtractive conceptions becomes a useful device for probing teachers' existing theories regarding their students, and how those theories interacted with district and school decisions regarding Proposition 227 to establish a context for classroom practice.

Competing Theories in Action: A District's Responses to 227

To understand how competing theories regarding the education of language minority students materialize into action, we examine select findings from one district's implementation of Proposition 227. Focusing on the responses of three teachers in the district, we examine how additive and subtractive theories influenced and shaped the nature of Proposition 227 implementation.

Walton Unified School District

Despite its attempt to prescribe a very uniform solution for the education of linguistically and culturally diverse students across the state of California, the law's impact on education of ELL students has varied widely from district to district, school to school, and in some cases classroom to classroom. Garcia & Curry-Rodriguez (2000) and Gandara et al. (2000) report that certain districts across the state have used the waiver clause of the law to pursue district wide waivers, others have implemented the English Only provisions of the law, and a third group has left the primary decisions up to individual schools. The implementation decisions made by "Westway" and "Open Valley," the two elementary schools which are the focus of this research, represent a microcosm of what occurred across the state. Each school took actions based upon "competing ideologies" about how schools should respond to the challenge of linguistic and cultural diversity.

Walton Unified School District devised a plan that maximized flexibility for Westway and Open Valley. Under the plan, the schools could choose between maintaining their bilingual programs through the parental waiver process, or developing a program for ELL students called "English Language Development" (ELD).

At Westway Elementary, all students who had been in bilingual programs were placed in self-contained ELD classes. The context for Proposition 227 implementation at Westway Elementary was shaped by the school's positive orientation towards English-only instruction and curricular control arrangements. The decision to shift to English-only was made by the school's veteran principal, 'Beverly Elmherst,' who in the past had tended away from hiring certified bilingual teachers. Consequently, the school had only three teachers who held the Bilingual Crosscultural Language and Academic Development (BLCAD) certificate. This hiring pattern meant there were very few strong advocates for maintaining the school's program.

The school's movement away from primary language use coincided with the state-wide decade-long movement toward phonics-based reading instruction and away from meaning-based or whole language instruction. A series of laws passed throughout the 1990s culminated in the California Reading Initiative (CRI), a collaborative effort between the state legislature, the Governor, and the California Department of Education. The new policy advocates a *balanced* approach to literacy instruction. It defines balanced literacy instruction with a definitive nod toward decoding and direct phonics instruction:

> A balanced approach involves considerable time and effort dedicated to basic decoding while attention is given to important meaning-based aspects of reading. For most students,however, intensive direct teaching of phonemic awareness, sound-symbol relationships, blending skills, and reading fluency is of primary importance. (CRI, p. 4)

The changes in literacy policy have positioned phonics and phonemic awareness as the primary concerns for early literacy instruction.

Consistent with the move on the state the level toward phonics-based instruction, in February of 1998, the school adopted Open Court Collections for Young Scholars (hereafter, Open Court) as the school wide language arts series. Open Court uses explicit teacher-directed instruction to teach phonemic awareness, phonics, and reading comprehension. During the instructional components of the program, which include teacher-directed writing and reading exercises, and skills practice drills, teachers use scripts for all teacher questions, prompts, and responses. During blending, a center piece of the program, teachers read all sounds of a word and have students repeat them. Reading and writing activities are tightly controlled by the teacher.

While the school-wide context at Westway was characterized by a lack of curricular freedom and a climate favorable to English-only, the local school context at Open Valley was quite different. First, the overall climate of the school showed an overwhelming commitment to the goals of bilingual and multicultural education. Second, teachers at Open Valley experienced a great deal of curricular freedom. In the fall of 1999, the teachers at the school mobilized to secure parental waivers in order to maintain the school's bilingual program. Nearly every child who was in a bilingual program prior to the proposition was in a bilingual program in the fall of 1999. To avoid a second year of the waiver process in the spring of 1999, the teachers and administration of Open Valley applied for and received Charter status. Under California law, Charter status gave the school curricular flexibility and freedom from the direct mandates of Proposition 227.

Teachers

The research on the implementation of Proposition 227 focused on four teachers—two at both Open Valley and Westway. At Open Valley, the research focused on two teachers, "Elisa" and "Angelica".

Angelica, a fifth year teacher, came to teaching through her involvement in a migrant education program as an undergraduate. Although she was only in her late-20s, she had taught Sunday School for 12 years. She credited her experience with the migrant education program and her work in Sunday School as having a large influence on her teaching. She was born in Mexico, but attended school in California when her parents immigrated. During the year of the study, Angelica taught a 2nd grade bilingual class of approximately 18 students. The second grade students in her class received language arts and math instruction in Spanish. Instruction in the afternoon, which included art, ESL, and social studies, occurred in English.

Born in Mexico, Elisa was educated in California and grew up in the Central Valley. She had been a teacher for four years—all of them at Open Valley and each in a different grade. During the 1999-2000 academic year, Elisa taught a 3rd grade bilingual classroom of approximately 14-20 students. Elisa's decision to enter teaching was closely related to her experiences as a child. Elisa had worked in the fields of the Central Valley, and felt that experience helped her to identify with the instructional and social needs of her immigrant students.

Two teachers, "Celia" and "Connie", were the focus of the research at Westway, but in this paper we present findings only related to Connie. We have chosen to focus on Connie to examine the manner in which existing deficit orientations in teachers interact with a subtractive policy context. Additionally, Celia's rather complex and multifaceted reaction to Proposition 227 implementation has been examined in another article (Stritikus, in press).

Connie, a Portuguese-American with 11 years of teaching experience, had always been assigned a bilingual classroom but never remembers requesting to be a bilingual teacher. Because the structure of the bilingual program prior to Proposition 227 placed native language instructional responsibility in the hands of teaching aides, Connie never worked directly with her immigrant students in the area of primary language instruction. During the study, Connie taught a 3rd grade, self-contained *English Language Development* class of 20 students.

Data Collection and Analysis

The research began in the spring of 1998-1999 and continued the research through the end of the 1999-2000 academic year. The research took place at Westway Elementary and Open Valley Elementary, the two largest schools in the small rural district of Walton Unified.

Stritikus used multiple sources of data to build a picture of the implementation of Proposition 227, and observed the teachers in a host of different situations including classroom literacy instruction, grade-level meetings, all-school meetings, and district-level meetings concerning ELL issues. In addition, each teacher and other key participants in the district were interviewed.

Stritikus observed each of the teachers' classrooms a minimum of 21 times. During classroom observations, Stritikusfocused on the nature of literacy instruction. Scratch-notes (Emerson, et al., 1995) and audio recordings from observations were used to create detailed fieldnotes. After Stritikus left the research site in May 2000, he completed a close reading of the entire set of fieldnotes looking for "certain words, phrases, patterns of behavior, subjects' ways of thinking, and events that stand out" (Bogdan & Biklen, 1992: 166). During these initial read-throughs, he asked questions of the data which centered on building an understanding of how teachers' beliefs and theories influenced literacy instruction in their post-Proposition 227 classrooms (Emerson, et al., 1995; Glasser & Strauss, 1967). To address each question, analytic commentaries grounded in the data were written. The analytic commentaries served as the basis of the themes generated from the data and were central in the development of codes and data analysis.

Findings

The manner in which the three teachers responded and reacted to Proposition 227 is illustrative of the way that subtractive and additive theories compete to shape the nature of the policy to practice connection. In large part, teachers guiding theories about their students influenced the way they mediated and negotiated the policy shifts brought about by changes in bilingual education policy. In the following sections, we explicate the connection between classroom practice and policy shifts by examining the role that teachers' theories played in the process. We highlight how aspects of a subtractive policy context brought certain aspects of teachers' additive or subtractive theories to the surface in their decision making process.

Teachers' Theories in Programs that Retained Bilingual Education

For Angelica and Elisa remarkably similar guiding theories drove their intellectual work at the school. Each teacher believed that native language instruction provided significant academic, cognitive, social, and cultural benefits for their students. For both teaches, the academic and cultural benefits of bilingualism were inextricable linked and strengthened their resolve and commitment to bilingual education. Angelica described her theory on the manner in which language minority students could most attain academic success:

> Yes, it is hard to remain a bilingual. But, if you don't give students a base—the foundation that the child needs to use against the second language—success in the second language is not going to happen. Basically, we are all here in this America. And, we do all need to speak the language of this country, but that doesn't mean that we have to let go of our language. (Angelica, Interview, May, 2000)

For Angelica, academic success and participation in American society did not mean that students had to sacrifice elements of their social and cultural identities. For her, these identities served as the basis for student success. Angelica believed that Proposition 227, and its supporters, were asking Latino student to leave crucial elements of their culture and language behind. She saw her role as a teacher to ensure that this didn't happen at Open Valley.

Elisa echoed many of Angelica's theories about the benefits of bilingualism and native language instruction. In addition, she saw the use of native language instruction as direct part of a strategy of the advancement of Latino students. The genesis of this theory was related to her own experiences as a migrant farm worker:

> The sun was coming out at five o'clock in the morning. I was there alone. There was nobody in the field. I was just left there and I was waiting for the people to get there. I was maybe fourteen or fifteen. I kept thinking: What I am going to do? I didn't want to work in the fields for the rest of my life. That's actually what brought me up wanting to teach. It's like: I want to do something productive for my people—for the kids and parents who work in the fields because I saw how hard they work and they really didn't make any money. So, I wanted to make a difference. That's why I became a bilingual teacher. (Elisa, Interview, April, 2000)

Both teachers' guiding theories saw bilingualism as a social and academic resource and viewed tapping into students' existing linguistic capacity as the best way to ensure their academic success. The teachers possessed theories which allowed for student to be multilingual and still play valuable and meaningful roles in U.S. society. For both teachers, these theories had their roots in their personal experiences and the benefits of bilingualism they had gained as well as hardship they experienced in schools which did not value their linguistic diversity.

From Theory to Action: Teaching in a Bilingual School

For Angelica and Elisa, theories about language minority students lead to particular types of responses to policy shifts. Angelica, for example, became a very vocal proponent for bilingual education after the passage of Proposition 227. She used her standing in the school to rally support for the school's bilingual program and helped secure the parental waivers necessary to continue bilingual education at the school. Each teacher used native language instruction in real and substantial ways in their classroom, which included assessments done in English and Spanish. Both teachers commented that Proposition 227 had renewed their commitments to bilingual education. Angelica continually looked for opportunities to defend the school's program and petitioned the district for resources related to bilingual education.

For Elisa, her renewed commitment was directly related to the manner in which she saw language use in her classroom:

> Creo que me hizo un poco mas rebelde [I think it has made me bit more rebellious] about using Spanish. Before I was like I shouldn't speak in Spanish, because we are being asked to move away from bilingual education. But, now, I don't feel that way. (Elisa, Interview, April, 2000)

Her commentary illustrates the manner in which the subtractive policy context brought certain elements of teachers' theories to the surface as they negotiated aspects of Proposition 227.

To understand the manner in which teachers' theories serve to mediate their responses to policy shifts, we present the following data excerpt from the first day of English-guiding reading groups in Elisa's third grade classroom. The event illustrates her attempt to create an additive context for learning in her classroom.

Elisa commented that the debate over Proposition 227 had made her more committed to making sure that her the students saw their home language as a resource. On the first day of English guided reading, Elisa lead each of the groups that she worked with through a series of activities in Spanish. Each of the five groups she worked with examined a picture of animal. Elisa solicited comments from the students about the animal. After having a conversation about the picture, Elisa gave the students five minutes to write a few sentences about the picture. The following conversation occurred between Elisa and Ernesto, Rosa, Cristóbal, Betty, and Daniel after they had concluded the activity.

Elisa	Students	Action
Ok, ayúdame. Qué es lo que estamos haciendo? (2) Por qué estamos haciendo esto?		
	Ernesto: (dutifully) Aprender.	
	Rosa: Para aprender los dibujos.	All the students have raised their hands and she is calling on them by touching her hand in front of the students.
	Cristóbal: Para aprender más palabras.	

Para aprender palabras?		
	Rosa: Para hacer como agarrar palabras de un dibujo.	
[E repeats Rosa's response.] Y para... lo que dijiste ahorita. Si ven como acerlo en español. Y cómo vamos aprender el inglés.	Cristóbal: (take turn with out hand up) Para aprender el inglés.	
	Ernesto: Aprendiendo palabras.	
Aprendiendo palabras.		
	Elsa: (Hand up--Officially recognized) Tenemos que saber como lo hacemos en español primero y luego es más fácil hacerlo en inglés.	
Si cuando estamos con un dibujo, y tenemos palabras, y de las palabras qué hacemos?	Daniel: Oraciones!	
De estas oraciones qué podemos hacer?	Ss: Párrafo.	
De un párrafo qué podemos hacer?		At this point the pace of the discussion quickens.
	Ss: Un capítulo Un resumen Ensayos.	
Qué tiene que ver esto con el inglés?		
	Daniel: Yo voy a saber las palabras que tiene que responder.	
	Betty: Puedes poner tree en vez de árbol (She is pointing at the white board where Elisa had written some of the sentences students had generated.)	After Betty's comments Elisa asks the students using "what can I put in place of.." with each of the Spanish words that they had come up with to describe the picture. The student excitedly call out the English words.
Si saben las palabras en inglés, podemos hacer oraciones en inglés.	Ss: Sí.	
Y luego podemos hacer párrafos.	Ss: Sí.	
Luego podemos hacer ensayos en inglés.		

After this exchange, Elisa tells the students that when they are learning a second language their mind will have to work extra hard, and that sometimes they will have to think first in Spanish to get the job done.

English Translation

Elisa	Students	Action
Ok, help me out. What are we doing? Why are we doing this?		
	Ernesto: (dutifully) To learn. Rosa: To learn about drawings. Cristóbal: To learn words.	All the students have raised their hands and she is calling on them by touching her hand in front of the students.

To learn words?		
	Rosa: To learn how to take words from a picture.	
[E repeats Rosa's response.] Y for… What do you just say? You can see how we do it in Spanish Y how are we going to learn English?	Cristóbal: (take turn w/o hand up) To learn English.	
	Ernesto: Learning words.	
Learning words.		
	Elsa: (Hand up--Officially recognized) We have to know how to do it in Spanish first and then it will be easier to do it in English.	
Yes, and when we are working with a picture, and we have words, for the words what do we make?	Daniel: Sentences	
And from those sentences what can you do?	Ss: Paragraphs.	
From one paragraph what can we do?		At this point the pace of the discussion quickens.
	Ss: A chapter A summary An essay	
What does all this have to do with English?		
	Daniel: I'll know the words I need to know to answer the questions.	
	Betty: You can put 'tree' [said in English.] in place of arbol. (she is pointing at the white board where Elisa had written some of the sentences students had generated.)	After Betty's comments Elisa asks the students using "que puedo poner en vez" with each of the Spanish words that they had come up with to describe the picture. The student excitedly call out the words.
If you know English, can we do sentences in English?	Ss: Yes.	
And later can we do paragraphs?	Ss: Yes.	
And later can we do essays?	Ss: Yes.	

The nature of teacher and student interaction on the first day of English guided reading was very telling. The message of the exchange was clear: "If you can do it in Spanish, you can do it English." Students were eager participants in these types of conversations and shared stories about bilingual relatives or about community members who spoke English and Spanish fluently

Elisa's decision to establish an instructional context in which Spanish was presented to the students as a direct way to make sense of English also had important consequences in terms of the way students approached learning tasks in the guided reading group. During the interaction of this group, the students eagerly explored the new ways they would be able to use English. Her framing of learning English as an activity created a sense of excited energy for the students. This excitement surfaced as the students discussed what they would one day be able to do with English. Daniel proclaims that he "will know the words that he has to know to respond [to questions]." And, Betty unsolicited offers her English knowledge to the group suggesting that Elisa substitute the tree for Spanish word "arbol." Elisa created an additive context in which she encouraged students to capitalize on their existing linguistic resources during their acquisition of English. The context established by Elisa made it clear to the students that Spanish was viewed as a language learning resource by their teacher. Because the focus of the study was to understand teachers' conceptions of bilingualism, we can not with certainty claim that this additive conception had a direct impact on students' conceptions of their own bilingualism. For Elisa, however, her additive conceptions of bilingualism had influenced the manner in which she reacted to the subtractive policy context created by Proposition 227.

Angelica: Bring the Fight to Her Classroom

Angelica's commitment to bilingual education stemmed from the instances of racism she experienced as a child in American schools. She believed that her language and culture had been "devalued" by her own school experience. The nature of literacy instruction in her classroom seemed to be a direct response to her experience and the subtractive logic behind Proposition 227. Angelica's saw Proposition 227 as a direct challenge to her ability to provide an additive education for her Latino students:

> When Proposition 227 was happening: I saw a lot of Mexicanos on the news that said: Yeah, we live in this country and we have to speak English. Who are you to speak for somebody else? That's your opinion and maybe you don't want to be bilingual and maybe you fell in love with culture and left your past behind. But, there are many of us who don't want to leave our past behind (Angelica, Interview, May 2000).

Angelica strongly rejected the assimilationist implications of Proposition 227, and saw her classroom as the place to begin to counter the influence of the new law.

In the area of literacy instruction, Angelica believed that reading comprehension would improve if students were able to "see themselves in the story." In her perpetual "fight" for her students, Angelica learned a great deal about the lives of her students. She learned about their families, their siblings, and their home environments. Her knowledge of the students' social worlds was manifested in her interaction with them. Angelica used the knowledge that she had of the students to help them negotiate the stories they read. In helping the students access this knowledge in their negotiations of the written texts, she used a questioning strategy that facilitated student participation (Garcia, 1999; Jimenez & Gersten, 1999). These strategies were evident in the following literacy event as Angelica had a pre-reading discussion with a small reading group.

> On the first day of her work with the group, she had given the students extended turns to talk about their experience with pets in their homes and neighborhoods. Over the course of reading the story, Angelica frequently drew upon those discussions in her questions about the story. During this event, Angelica integrated knowledge about her students directly into the instructional discourse.
>
> Six boys are seated at the reading table. The boys are about to read the story "Enrique y Pancho."
>
> Angelica tells the students that they are going to read a story about a narrator who is an only child, and that she wants them to think about what life might be like for an only child.

Angelica	Students	Action
En un a familia pequeña juegas solo, ¿que tienes que hacer en una familia grande? Cesar.		As Angelica is asking the question, Cesar has pushed his chair three feet away from the reading table. He is looking around the room with a blank stare.
	Cesar: Eh, Yo no sé.	
(Sternly) Piensa. Si en una familia pequeña juegas solo-- que tú nada más fueras el único niño en tu casa-- Tú tienes que jugar con tus juguetes solito, ¿verdad? (2)		
	Cesar: (looking up with a sort of "you caught me grin") ¿ Si alguien estaba?	
Y si está Anna (his sister) y todos tus hermanitos, ¿ qué debes hacer con tus juguetes?		
	Cesar: ¿Eh?	
¿Qué se debe hacer con tus juguetes si están todos tus hermanitos en tu casa?		
	Cesar: Juegan.	Juan who raised his hand after Angelica's initial questions has a strained look on his face--as if he can't wait to participate. His hand was initially raised, but he lowers it as Angelica talks
¿Tú juegas con quién?		
	Cesar: Con mi hermano.	

¿No más con él o con todos?		to Cesar. His eager facial expression never diminishes.
	Cesar: Con todos…	
A ver, dime en una oración.		
	(Continuing off of his last comment) …Ya no, mi hermano y Anna no pueden jugar juntos, y este	
OK, pero Cesar. Si tú juegas solo en tu casa, y como tú tienes una familia grande, qué debes hacer. ¿Compartir o jugar solo?		
	Cesar: Compartir con otros. (pauses and smiles) Tengo que compartir mis juguetes.	
A: Muy bien, ¿están de acuerdo con Cesar?		
	Other boys: (With enthusiasm) Sí!	
	Juan: (speaking out without being formally acknowledge by Angelica) Iba a decir compar compar (Excitedly to the rest of the group). Es lo que iba a decir. Juego con todos.	
		The other boys at the table nod eagerly at Juan.

In the moments that followed this interaction, the students were told to open their reading books. When they did, they read with the enthusiasm and expression of a stage performance.

English Translation

Angelica	Students	Action
In a small family you play alone, what do you have to do in a big family? Cesar.		As Angelica is asking the question, Cesar has pushed his chair three feet away from the reading table. He is looking around the room with a blank stare.
	Cesar: Eh, (quickly and chewing his words) I don't know.	
(Sternly) Think, if in a small family you play alone—and you are the only kid in your house—You have to play alone, right?		
	Cesar: (looking up with a sort of "you caught me grin") If some one is there?	
And if Anna is there and all your brother and sisters, what do you have to do with your toys?		
	Cesar: Eh?	
What do you have to do with your toys if all your brothers and sisters are in the house?		
	Cesar: They play.	Juan who raised his hand after Angelica's initial questions has a strained look on his face--as if he can't wait to participate. His hand was initially raised, but he lowers it as Angelica talks
Who do you play with?		

	Cesar: With my brother.	to Cesar. His eager facial expression never diminishes.
Only with him or with the others?		
	Cesar: With the others…	
OK, give me a sentence.		
	(Continuing off of his last comment) …Now no, my brother and Anna can't play together because…	
OK, but Cesar. If you play alone in the house, and how you have a large family, what do you have to do? Share or play alone?		
	Cesar: Share with the others. (pauses and smiles) I have to share my toys.	
A: Very good do you agree with Cesar?		
	Other boys: (With enthusiasm) Yes	
A: I am going to write plays with others.	Juan: (speaking out without being formally acknowledge by Angelica) I was going to say sha shar Juan:	
	(Excitedly to the rest of the group). That is what I was going to so. I play with everyone.	A writes "You play with the others" in the column under "A large family."

This exchange and the questioning strategy used by Angelica demonstrated the place that students' home culture had in shaping the nature of literacy instruction. By capitalizing on her knowledge of Cesar's home life, Angelica gave him a way to be a meaningful participant in a discussion that he had otherwise started to ignore. The knowledge she had of his family had come from the many visits she had made to his home. Her knowledge of his social and cultural world served as an instructional life preserver allowing Cesar to construct a response to a question to which he was struggling to respond. Literacy research has long documented that making space in the official curriculum for the lives of students opens up new avenues for students' learning (Dyson, 1993). Angelica's decisions in literacy instruction represented her commitments to creating a meaning-centered literacy learning context for her students. Garcia (1999) and Jimenez and Gersten (1999) have documented that such an approach is essential in the literacy development of Latino children.

In addition, the nature and shape of the discussion had the effect of keeping all students engaged—even Juan—who although he was not able to participate in the direct exchange was still able to share that "this is what I was going to say." Angelica set the stage for this exchange by informing the boys that were having this discussion for a particular reason—a reason directly connected to their ability to relate to the story they were about to read.

Additive Theories Summary

Elisa and Angelica took steps in and out of their classroom to limit the impact of Proposition 227 on their school's bilingual program. Angelica was a central player in securing parental waivers necessary to maintain the school's bilingual program. Elisa became "mas rebelde" (more rebellious) and more purposeful in her attempts to bridge her students' Spanish and English literacy development. As a first grade teacher, Angelica was primarily responsible for Spanish literacy development. As the third grade teacher in the school's bilingual program, Elisa was primarily responsible for 'transitioning' the students from instruction primarily in Spanish to instruction primarily in English . Given the differences in their teaching situations, cross case comparisons are difficult. None the less, similar themes did emerge in the coding of teacher-run literacy from both classrooms. The coding of these events highlight the additive nature of the classroom contexts created by the two teacher. Both teachers created contexts which the following types of interactions were most prominent:

- *Events of Story Question.* Teacher questions soliciting retelling or summary of events from a story.
- *Creating/building on Intertextuality.* Comments or questions that drew upon students' social and cultural lives as resources in understanding stories the class read.
- *Concept Question.* Questions that asked students to draw conclusions or make inferences about events

or concepts in stories. During English instruction, teachers encouraged students to respond in Spanish if they were not able to do so in English.
- *Turn Extension.* Comments or questions by teachers that extended student turns.

Because we did not collect any data prior to the passage of Proposition 227, it is not clear if teachers demonstrated this type of literacy prior to its passage. Thus, Proposition 227 did not cause an additive orientation for Elisa and Angelica, but rather their experiences of the Post-227 context reinforced and refined their existing additive conception of their students and bilingualism.

Subtractive Theories of Education for Language Minority Students

To understand the connection between subtractive theories for language minority students and classroom practice, we present the case of Connie, a third grade teacher at Westway Elementary. Connie's case is illustrative of how teachers' existing subtractive theories materialize in classroom practice. Her case is instructive because her theory regarding the education of language minority students mirrored the theories of many school- and district- leaders who eliminated their bilingual education programs after the passage of Proposition 227 (Garcia & Curry-Rodriguez, 2000). Proposition 227 did not cause her subtractive orientation but rather reinforced it and gave her new opportunities to act upon it.

Connie's theories surrounding her students were undergirded by two major beliefs about the education of language minority students. First, she believed that the English language served as a unifying force in the United States that was undermined by multilingualism. In this sense, Connie was in striking agreement with much of the political discourse surrounding both the English-only and anti-bilingual education movements. In an interview, Connie commented, "I totally agree that English should be **the** language of this country. You need to have some base and I think English needs to be the base here." A child of Portuguese immigrants, Connie resented the "special treatment" that she felt Latino children and families received. She viewed bilingual education as one such "special treatment."

Second, Connie believed that her students' academic progress was severely limited by their use of Spanish. Thus, rather than seeing students' primary language as a resource, she saw it as one of their primary weaknesses:

> My students' problem is that they rely too much on their Spanish. I know a lot of them came from 2nd grade classes where they spoke Spanish all the time to the teacher. It makes a big difference. My goal is that they learn as much vocabulary as they can, learn to speak grammatically correct, and have their adjectives and nouns in the right places. (Connie, Interview, December, 1999)

Connie felt that the students' use of Spanish interfered with their acquisition of English. This subtractive theory differs shapely with the important theoretical work done in the area of second language acquisition stressing the transfer of academic and cognitive skills independent of language (Cummins, 1979).

From Subtractive Theory to Subtractive Practice

Connie's theory about language minority students resulted in a particular kind of educational practice which did not focus on the cultural, social, and linguistic resources brought by her students. Watered-down and deficit-based literacy practice in the new policy environment reflected Connie's instructional goals and expectations for her students (Gersten & Woodward, 1992; Ramirez, 1992). A significant amount of instructional time focused on phonetic exactness—moments in instruction when Connie focused on the components and sounds of words. During these interactions Connie's emphasis was on correct pronunciation and strict adherence to following directions. Coding of literacy events revealed that Connie's literacy practice centered on the following types of interactions:

- *Word Meaning.* Connie asked the students about the meaning of an individual word. She used the word in a sentence until the students could supply a synonym.
- *Conventions.* Connie asked the students about the punctuation of a particular sentence, or she asked students to identify words that were particular parts of speech in the text.
- *Phonetic Exactness.* Connie worked with the students to ensure the proper pronunciation of English words and phonemes.

Her emphasis on these three types of interactions was influenced by the nature of the Open Court program and its literacy material. During teacher run reading events, Connie seldom asked questions regarding the story events or the plot. Connie often asked students to identify compound words or to circle long vowels. Such interaction contributed to the treatment of text as a puzzle. Texts were viewed as little more than the sum total of their phonetic or grammatical values. During literacy instruction, Connie closely adhered to the script of the Open Court teacher's manual. Open Court activities dominated her instructional day. Beyond the 40 minutes that Connie spent in math instruction, the entire day was occupied with Open Court literacy activities.

The following literacy event, highlights the nature of literacy instruction in Connie's classroom and the manner that Connie's beliefs about her students, which were in large part influenced by her own familiar experience, seemed to influence the enactment of such literacy practice.

> Connie stood at the front of the class and had just read the first problem of the worksheet. She instructed students that they were supposed to circle each long vowel sound in each of the sentences and write the word in the long vowel column. This was the third in a series of worksheets the class had done that day. Connie completed the first three sentences with the students. In each sentence, her pattern was fairly consistent. She read the sentence and asked the students which words in the sentence had a long vowel sound. Students were not allowed to pick up their pencils until the class had identified all the long vowel sounds. During the first three sentences, a few student called answers without being officially recognized. When this happened on the 4th sentence, Connie said, "Since you seem to have no problem with this activity you can do it on your own."
>
> Ruben and Miguel, who were seated on the opposite side of the room from where I was, excitedly rubbed their hands together. I got up from my seat and sat behind Miguel, a child who always seemed to have a smile on his face.
>
> Miguel: (Reading number 5) (Reads in a flat tone with no questioning intonation.) Will Pat go to the store. (Pauses for a moment) Will Pat **go to** the store. (Flat intonation). Will Pat...Pat go to the store? (An almost raised but unnatural intonation on store). [He raises his head from the text]. That doesn't make any sense. (almost smugly) Don't matter. [He picks up his pencil and writes the words "go" and "store" in the Long O column.]

This literacy event highlighted many of the themes which emerged from the study of Connie's classroom. Classroom instruction focused on the component parts of reading. Connie's comfort with this focus was related to her views about the instructional needs of her students. The event also highlighted the tightly controlled nature of literacy events. In the activity—as was the case with many others—students were allowed to do the work independently only as a form of punishment. Lastly, the event indicated the nature of students' experiences of literacy curriculum which stressed skills over meaning.

Connie believed that her students would experience success if they stopped speaking Spanish in the classroom. During grade level teacher meetings, Connie voiced this position. Her comments generally related to "deficits" in the students (Lipman, 1998). While it is highly likely that Connie's deficit perspectives of her students existed prior to Proposition 227, she noted that Proposition 227 had allowed her to act on her beliefs about the needs of her students in ways that she had not been able to. Because she was convinced that several issues outside the realm of her classroom contributed to the academic failure of the students, she took no actions to the change the programmatic and curricular actions at the school.

Connie's ideological alignment with the law and her views about her students played a large role in the connection between policy and practice in her classroom. The following literacy exchange taken from a field note entry occurred early in the year and was indicative of her priorities and perceptions:

> Connie told the class that she wants them to work on their reading comprehension, and that to do so they are going to read stories on worksheets that will help them understand other stories better. Connie told the students to place their fingers under the first word and called on individual students to read a story about a snow flake. The story was a part of the first grade skills practice of Open Court. She called on Sonia, who struggles to read the first sentence of the eight sentence story.
>
> Connie said, "OK, Sonia, since Luis is ready to read I am going to give him a chance." Luis, a recent immigrant in the class with very strong decoding skills did not understand Connie's request as a request to read because he wasn't called on directly. He stared at Sonia and then turned his gaze back to the teacher. Connie nodded at him, and he still looked confused. A student sitting next to him said in a quiet voice, "Tienes que leerlo" [You have to read it.] Connie clinches her fist, "Uhg," she said with great exasperation, "Don't say it in Spanish!" (October 21, 2000).

The event which was similar to many literacy events in Connie's classrooms speaks to two beliefs that guided her approach to classroom literacy instruction: 1) Spanish was a detriment to her students' academic progress, and 2) what her students needed most were the "basics." Connie's interaction with the local school context influenced the way these beliefs surfaced in her classroom literacy practice. She noted that the move to English-only made her feel more comfortable in stopping her students from speaking Spanish in the classroom.

Subtractive Summary

For Connie, Proposition 227 offered an opportunity to enact a subtractive version of language and literacy practice in her classroom. Literacy instruction in her classroom was heavily influenced by her theories about her students and their bilingualism. Proposition 227 and its subtractive implications for the schooling of culturally and linguistically diverse students complemented Connie's existing views of her students and gave her liberty to attempt to restrict and limit students' use of Spanish in her classroom. While we do not claim that Connie's use of the Open Court literacy series is representative of all uses of the program, Connie's case illustrates how teachers with subtractive theories of their students might utilize and implement aspects of similar skills-based scripted literacy programs.

Conclusion

California has begun a weighty experiment in the instruction of language minority students based upon subtractive theories of education. The underlying theory of Proposition 227 suggests that linguistic diversity is a problem in need of correction, and instruction exclusively in English provides the best therapy for such deficiencies. Such a theory of instruction suggests that the primary role of schooling is Americanization.

Proposition 227 is not just a theory, but one of the dominating policy voices in California and the nation guiding the schooling of linguistically diverse students. Given that teachers will continue to be the last line of implementation in this growing policy trend, it is important consider various aspects of the roles they play. In this article, we have chosen to focus on the role that teachers' existing theories about their students play in the way they reacted to aspects of Proposition 227. To understand the range of teachers' theories, we have presented distinctions between additive and subtractive conceptions of language minority students.

The distinction between additive and subtractive conceptions of schooling for culturally and linguistically diverse students are a useful tool for understanding how teachers' existing theories were complemented or contrasted by Proposition 227 implementation. For Angelica and Elisa, the two teachers with strong additive conceptions of schooling, Proposition 227 served to strengthen their commitments to bilingual education. Their renewed commitments were evident in the manner in which they framed their classroom practice in relation to Proposition 227. Elisa asserted that Proposition 227 had made her more determined to use students primary language as a resource; and, Angelica, compared Proposition 227 to getting knocked down in a soccer game:

> Proposition 227 really pushed the people that did believe in continuing fighting for our dream. It's like a soccer game. You didn't make the goal. Oh, well. You have a chance of getting up and trying again. Soccer players fall many times during a game. They trip over each other. We can trip over these polices and fall over these laws. You can trip me, and I'll fall, but I'm going to get up again. I'll keep going. When things like Proposition 227 happen, just don't trip, fall, and stay laying down. (Angelica, Interview, May, 2000)

Angelica saw Proposition 227 as one major impediment to enacting an additive conception of education, but she saw it as a challenge she could and would overcome.

The additive conceptions possessed by Elisa and Angelica served as a basis for how they reacted to and mediated aspects of Proposition 227 implementation. Angelica was a key member in securing parental waivers at the school which enabled the school to maintain its bilingual program. Both teachers saw the manner in which they constructed their classroom literacy practice as a response to the political and pedagogical implications of Proposition 227 implementation. They considered their attempts to create an additive classroom context for their students as ways to fight for bilingual education. Thus, the teachers' additive conceptions played a significant role in the actions the two teachers took in and out of their classroom contexts.

For Connie, her subtractive conceptions of her students were complemented by the political and pedagogical implications of Proposition 227. In many senses, the subtractive policy context served to clarify her pedagogical purpose. A she interacted with the local policy context, the result was an enactment of practice which was a direct match of the intent of the new law. As her students became more resistant and distant based upon their experience of Open Court, she became more convinced that her students needed more "basic" instruction. As her experience as part of family that "made it" without any special programs influenced her views of the policy, she become more convinced that any Spanish usage in her class was detrimental to student learning. She enacted punitive rules for students who used Spanish and noted that Proposition 227 had given her the feeling that this was a proper course of action to take.

The importance of teachers' theories and beliefs has been supported by Wiese's (2001) examinations of a policy and practice at a dual language immersion school after the passage of Proposition 227. She found that the manner in which Federal and State educational policy is reconstructed at school level was highly influenced by teachers' theories about their students, instruction, and the world around them.

Seeking the day when all language minority students will conclude that what they do in their classrooms does matter, we suggest that the theories that teachers hold about their students and instruction play a monumental role in the face of educational polices designed to lead to specific practices. Theories can bolster the intent of the policy, as was the case with Connie and the teachers at Westway Elementary, or theories can provide teachers with a powerful basis to resist and reshape the intended consequences of certain policies. If teachers are to capitalize on the linguistic, cognitive, and cultural resources which language minority students bring to the classroom, then those concerned with education must continue to pursue and develop substantial ways to support and develop additive conceptions of linguistic diversity in teachers.

References

Auerbach, E. (1995). The politics of the ESL classroom: Issues of power in pedagogical choices. In J. Tollefson (Ed.), *Power and inequality in language education*. New York: Cambridge Press. 9-33.

Banks, J.A. (1995). Multicultural education: Historical development, dimensions, and practice. In J. A. Banks, & C. A. McGee-Banks (Eds.), *Handbook of research on multicultural education* (pp 3-34). New York: Macmillan.

Bogdan, R., & Biklen, S. (1992). *Qualitative research for education: An introduction to theory and qualitative methodology*. Needham Heights: Allyn and Bacon.

Cohen, D. K. & Barnes, C. A. (1993). Pedagogy and policy. In D. K. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Issues for policy and practice* (pp. 207- 239). San Francisco: Jossey Bass.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research 34*, 46-49.

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Great Britain: Multilingual Matters.

Dyson, A. Haas. (1993). *Negotiating the permeable curriculum: On the interplay between teacher's and children's worlds*. Urbana, IL: NCTE

Emerson, R., Fretz, R., & Shaw, L. (1995). *Writing ethnographic fieldnotes*. Chicago: The University of Chicago Press.

Gandàra, P., Maxwell-Jolly, J., Garcia, E., Asato, J., Gutiérrez, K., Stritikus, T., & Curry, J. (2000). *The Initial Impact of Proposition 227 on the Instruction of English Learners*. Santa Barbara, CA: Linguistic Minority Research Institute. Available on the www at uclmrinet.ucsb.edu.

Garcia, E. E. (1995). Educating Mexican American students: Past treatments and recent developments in theory, research, policy and practice. In J. Banks & C. A. Banks (Eds.), *Handbook on research on multicultural education* (pp. 372-384). New York: MacMillan.

Garcia, E. (1999). *Understanding and meeting the challenge of student cultural diversity*. Boston, MA: Houghton Mifflin Company.

Garcia, E. and Curry-Rodriguez, J. (2000) The education of limited English proficient students in California schools: An assessment of the influence of Proposition 227 on selected districts and schools. *Bilingual Research Journal* 24 (1 & 2), 15-35.

Garcia, E. & Gonzalez, R. (1995). Issues in systemic reform for culturally and linguistically diverse students. *College Record, 96*(3), 418-31.

Gersten, R. M. and Woodward, J. (1992) The quest to translate research into classroom practice: Strategies for assisting classroom teachers' work with "at risk" students and students with disabilities. In D. Camine and E. Kameenui (eds) *Higher Cognitive Functioning for all Students*. (pp. 201-218). Austin: Pro-Ed.

Glasser, B.G. & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Gutiérrez, K., Baquedano-Lopez, P., & Asato, J. (2000). "English for the children": The new literacy of the old world order, language policy and educational reform. *The Bilingual Research Journal, 24* (1 & 2), 87-105.

Heller, M. (1994). *Crosswords: Language education and ethnicity in French Ontario*. Berlin and New York: Mouton de Gruyter.

Jimenez, R., & Gersten, R. (1999). Lessons and dilemmas derived from the literacy instruction of two Latina/o Teachers. *American Educational Research Journal 36* (2), 265-301.

Kennedy, M. (1991). *An agenda for research on teacher learning*. East Lancing : Michigan State University, National Center for Research on Teacher Learning.

Kerper-Mora, J. (2000). Policy shifts in language-minority education: A misma between politics and pedagogy. *The Educational Forum, 64*, 204-214.

Lipman, P. (1998). *Race, class, and power in school restructuring*. Albany: SUNY Press.

Moll, L.C. (ed) (1994). *Vygotsky and education: Instructional implications and applications of sociohistorical psychology*. Cambridge: Cambridge University Press.

Olneck, M. R. (1995). Immigrants and education. In J. A. Banks, & C. A. McGee-Banks (Eds.), *Handbook of research on multicultural education*, (pp 310-327). New York: Macmillan.

Olsen, L. (1997). Made in America: Immigrant students in our public schools. New York: The New Press.

Ramirez, J.D. (1992) Executive summary. *Bilingual Research Journal* 16 (1& 2), 1-62.

Sekhon, N. (1999). A birthright rearticulated: The politics of bilingual education. *The New York University Law Review, 74* (5), 1407-1445.

Stritikus, T. (2002). *Immigrant children and the politics of English-Only: Views from the classroom*. New York, NY: LFB Scholarly Publishing.

Stritikus, T. (in press). The Interrelationship of Beliefs, Context, and Learning: The Case of a Teacher Reacting to Language Policy. *Journal of Language, Identity, and Education*.

Wiese, A.-M. (2001). *"To meet the needs of the kids, not the program": Teachers constructing policy, program, and practice in a bilingual school* (Dissertation). Berkeley: University of California, Berkeley.

Wiese, A. & Garcia, E. (1998). The bilingual education act: Language minority students and equal educational opportunity. *Bilingual Research Journal, 22*(1), 1-18.

Wong Fillmore, L. (1992). Against our best interests: The attempt to sabotage bilingual education. In (Ed.) J. Crawford, *Language loyalties: A source book of the official English controversy*. Chicago: University of Chicago Press.

Woods, P. (1994) Adaptation and self-determination in English primary schools. *Oxford Review of Education* 20(4).

About the Authors

Tom Stritikus
University of Washington
College of Education
122 Miller Hall
Box 353600
Seattle WA 98119
Email: tstrit@u.washington.edu

Dr. Tom Stritikus is Assistant Professor, in the College of Education, University of Washington. He earned his Ph.D. in 2000 from the University of California, Berkeley. He is a faculty associate of the Center for Multicultural Education at the University of Washington. His teaching and research focuses on policy and practice issues for culturally and linguistically diverse students.

Eugene Garcia
Arizona State University
College of Education
PO Box 870211
Tempe AZ 85287-0211
Email: Eugene.garcia@asu.edu

Dr. Eugene Garcia is Vice President for University-School Partnerships and Dean of the College Of Education at Arizona State University. He received his B.A. from the University of Utah in Psychology and his Ph.D. in Human Development from the University of Kansas. He has served as a Post-Doctoral Fellow in Human Development at Harvard University and as a National Research Council Fellow. He has been a recipient of a National Kellogg Leadership Fellowship and has received numerous academic and public honors.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 1494 times since August 7, 2003

Volume 11 Number 27 **August 7, 2003** **ISSN 1068-2341**

Research or "Cheerleading"?
Scholarship on Community School District 2, New York City

Lois Weiner
New Jersey City University

Citation: Weiner, L. (2003, August 7). Research or "cheerleading"? Scholarship on Community School District 2, New York City. *Education Policy Analysis Archives, 11*(27). Retrieved [date] from http://epaa.asu.edu/epaa/v11n27/.

Related article:
Vol. 11 No. 28

Abstract

This article examines data on student achievement and school demographics not explored by the researchers who have promoted Community School District 2 (CSD 2) as a model of urban school reform that should be replicated elsewhere. Data on achievement indicate a remarkable degree of social and racial stratification among CSD 2's schools and levels of achievement that closely correlate with race, ethnicity, and poverty. In addition, when CSD 2's scores on state and city tests of mathematics are compared with results from CSD 25 in Queens, a school district that serves a population demographically similar, the superiority of its functioning becomes questionable. The article explains why the design of research on CSD 2 illustrates the perils to both research and policy when university-based researchers assume the role of "cheerleader" (Cuban, 1988), promoting reforms they have aided in implementing and assessing.

The direct assistance that university-based researchers provide to school systems involved in reform is generally accepted as positive, strengthening the relationship between theory and practice and in the process improving both. However, as Cuban (1988) notes in his "word to the wise" about researchers advising policymakers, "the legacy of disappointment with researchers who have been cheerleaders for this or that approach to be used in classrooms is not one to be envied" (p. 293). In this study I explain how reform in Community School District 2 in Manhattan and researchers' roles in advising, assessing, and promoting the reform model adopted by CSD 2 and teacher union officials illustrate the perils to education and research when researchers become "cheerleaders" for reforms. (Note 1)

Background and Significance of the Study

In debates about how to improve student achievement in urban school districts, Community School District 2's (CSD 2) strategy of using professional development to implement national standards has been put forward as exemplary, an illustration of public education's viability in the nation's cities (Elmore, 1999-2000; Elmore & Burney, 1997a; Elmore & Burney, 1997b; Elmore & Burney, 1999; Elmore & Burney, February1999). Research reports have heralded CSD 2's focus on a centralized system of professional development linked to national standards as being the key to improving achievement in urban schools (Fink & Resnick, 1999; Resnick & Harwell, 2000; Stein, D'Amico & Johnstone, April 1999). CSD 2's model is described by its most recent Superintendent as "delivering a world class education for every student through a redesigned labor management system that supports high performance learning communities utilizing the New Standards 'performance standards' along with city and state assessments" (Harwayne, 2000). Typical of the commendation of CSD 2 produced by researchers who have aided the district is this description:

> Over an eleven-year period, Community School District Two in New York City has amassed a strong record of successful school improvement in a very diverse urban school setting. Not only have test scores risen, but there is also a remarkable professional spirit among the teachers, principals, and central staff members of the district, which has 22,000 students in 45 schools (Fink & Resnick, 1999, p 3).

Publications ranging from the *TheWall Street Journal* to the monthly magazine of the American Federation of Teachers (AFT) have praised the model. In this article I scrutinize evidence for the claims that the model has been successful and that it can and should be replicated in other urban school systems. Scholarship on CSD 2 that promotes its success identifies the district and key personnel by name; my discussion adopts the same protocol. My rationale for naming the district and researchers is that commendations of the model have linked it explicitly to CSD 2 officials and have created a legitimacy, among politicians and in the media. The model has attracted attention even beyond urban school districts. One high-ranking state education official in Vermont pressed for CSD 2's curricular practices and professional development model to be adopted there (personal communication with Vermont college administrator, April 2001).

Sources of the Data for the Study

I examine reports of researchers who have worked with CSD 2 officials to design, implement, and assess the district's reforms. I also draw on publicly available data on school achievement and demographics published in *The New York Times*, and on the New York State Department of Education and New York City Board of Education websites . To my knowledge although these data are easily accessible, they have not been used heretofore to compare student achievement in CSD 2 with achievement in other districts with comparable demographics. Other data sources I use are personal correspondence with CSD 2 teachers, copies of memos sent to faculty by school administrators in CSD 2, an unpublished report on CSD 2's math curriculum produced by a group of parents and mathematics professors at New York University, and field notes following conversations with principals, teachers union officials and teachers in CSD 2 schools.

All teachers and school administrators employed in CSD 2 who spoke and corresponded with me were informed beforehand that I might use the information they provided in a published study. In each case the people I interviewed or who provided me with memos did so on the condition that they remain anonymous as the source of information. All cited fear of reprisals from supervisors as the reason for confidentiality. I gathered information from four administrators (in three schools), ten teachers (in three schools), and four people holding elected positions in the teachers union, the United Federation of Teachers (UFT).

My access to informants in CSD 2 was facilitated by contacts I made in the course of my participation as a parent-activist in the John Melser Charrette School, or as it is more commonly called, PS 3 (Weiner, 2002). PS 3 is an arts-based, alternative school started by parents, with a unique history that makes it simultaneously a "school of choice" and a regular zoned school for the neighborhood, Greenwich Village (Zuckerman, 2001). PS 3 is probably one of the "off the screen" schools that CSD 2 officials identify as "not working within the District #2 framework. While student achievement in some cases is fairly strong, the district leadership has concerns about the quality of instruction and or leadership in these schools" (D'Amico, van den Heuvel, & Harwell, 2000, p. 6). My initial examination of research on achievement in CSD 2 stemmed from my interest in understanding if PS 3 was considered "off the screen" and if so, why. According to published test scores, PS 3 maintained the same level of achievement as the other school serving Greenwich Village, and I wanted to know the source of CSD 2 officials' concerns about its instruction or leadership.

However, the focus and scope of my inquiry changed after my preliminary look at data on achievement in other schools in CSD 2 and New York City. The overwhelming presence of racial and social segregation in CSD 2 schools and the correlations between segregation and low achievement levels prompted me to examine the design and conduct of research about the district's success in boosting student achievement. Both investigations are discussed in this article. The first section deals with the context in which reforms and research were formulated and the reasons critical questions were ignored; the second segment examines data (presented in tables 1-5) that suggest why the CSD 2 model may not be as successful as it has been promoted as being, as well as reasons that it may not be replicable in most cities.

Insular, Self-referential Research Design

From the start, critical perspectives on CSD 2 were omitted from research on its system of professional development, as this description indicates:

> We were trying to figure out which people in the district should be interviewed and observed in order to understand how the district functioned. Someone started to diagram the way in which teachers were expected to learn from principals and professional developers and each other within their school, while at the same time principals were expected to learn from the Superintendent and Deputy and from each other how to be better at their instructional leadership job. Someone else said, "It's like those nesting dolls people like to bring back from their travels"—and the name was born. The image seems to work because the dolls are each independent, free-standing "people," yet they share a common form—and you can't decide which is the most "important" doll, the tiny one in the middle that establishes the shape for them all or the big one on the outside that encloses them all (Fink & Resnick, 1999: p. 6).

The study design supports an analogy (of nesting dolls) that seems not to be seen as simultaneously hierarchical and exclusively self-referential. The analogy is also remarkable for being static and decontextualized; relationships among the "dolls" are unaffected by "outside" influences such as parent feedback or critical perspectives that might be provided by teachers or principals who disagree with the superintendent and by other researchers, people who do not fit into the nest. Resnick observes that the doll in the middle is "tiny" and "the big one on the outside...encloses them all." One way to view the nested dolls as Resnick suggests is that they share a common form so it is not clear where the power to shape the relationship resides. But another view is certainly possible, that the outer doll shapes the configuration, and size and power diminish as one moves to the inside of the "nest." Support for this latter interpretation comes from Resnick's description of hierarchical power relations in CSD 2, that principals learn from the Superintendent and Deputy, and teachers from principals and professional developers and peers. (The addition of "peers" in this description is interesting because it does not fit the "nesting doll" analogy.) However, the research design did not address the possibility that differences in power and status among the nesting dolls corresponded to historically derived bureaucratic relations in urban school systems (Tyack, 1974) that have been identified by a considerable amount of research, for instance Freedman (1987) and Knapp (1995), as engendering teacher dissatisfaction and "burnout."

One result of the self-referential nature of the research design, clear from the "nesting doll" analogy that was adopted, was lack of attention to the controversies that have roiled in CSD 2, especially centering on the math curriculum that is mandated in every school. An unpublished report of a group of parents and professors organized to oppose the District's math curriculum, written by a mathematics professor at New York University, noted that parent and teacher dissatisfaction with the curriculum and its inflexible implementation were significant. My interviews with teachers confirmed the NYU report's finding that there was widespread fear that teachers would be disciplined if they supplemented the mandated curriculum with other materials to prepare students for city-wide tests. District officials refused to approve any orders for math workbooks to be used as supplements to the official materials. In some schools, teachers resorted to photocopying entire books for their classes, with tacit support of principals who turned a blind eye to the practice. (Note 2)

Despite its reputation as a powerful political force, the United Federation of Teachers had not positioned itself as an advocate for teachers, who expressed fears of being disciplined for not "toeing the line" with regard to CSD 2 curricular and instructional mandates. On the contrary, the UFT's parent organization, the American Federation of Teachers (AFT) promoted CSD 2's policies in the union's national magazine (*American Educator*, 1999-2000). Two "chapter chairs," (personal communication, May 2000), teachers who are elected to be the union representatives in the building, who hear teachers' problems and begin the grievance process, noted that they viewed the UFT leadership as reluctant to pursue teachers' complaints, formally or informally.

Clear evidence of teacher dissatisfaction with the UFT's stance towards CSD 2 surfaced in the Spring 2000 election for the UFT's district representative. Since creation of the community school districts in New York City as a response to the call for community control, the UFT has had a parallel organizational structure. Each community school district in New York has a union representative, a "district rep" who negotiates issues of local concern with the district administration. The "district rep" is elected by chapter chairs, but elections are almost always *pro forma* because of the UFT leadership's control of the union apparatus (Weiner, 1998). However, CSD 2 chapter chairs elected a teacher running against the "heir apparent" of the UFT leadership (personal communication, chapter chair in CSD 2, May 2000). The UFT's position was that though CSD 2's administration often had trouble "hearing what teachers had to say," nevertheless CSD 2 officials had to be supported; their model of reform was not only superior to others, it was the only one that could convince the public that city schools could be salvaged (private conversation with UFT President, Randi Weingarten, Oct. 2000). The vote for the CSD 2 "district rep" demonstrated that chapter chairs rejected the UFT leadership's stance toward CSD 2 officials. Their dissatisfaction certainly raises questions about the extent to which the CSD 2's strategy for labor-management relations will be supported by teachers elsewhere. Yet opposition among CSD 2 teachers to district policies and the UFT's support of them is absent in research on CSD 2.

Problems arising from the insularity of the researchers and of the design of the research itself appear in what was perhaps the key study. As Harwell, D'Amico, Stein, and Gatti (2000) note "A shortcoming shared by previous research done on the effectiveness of District #2's professional development system...was that the units of analysis used in these studies were schools. As a result, variation among students' performance and teachers' experiences within schools was ignored"(p.7). Hence their study attempts to correlate achievement on tests with teachers' professional development by examining test scores of individual students taught by teachers who described their professional development experiences in questionnaires. However, the questionnaire required teachers to provide their New York City Board of Education ID or "file" numbers (often used in lieu of names to locate personnel in official records) and their schools.

Readers of HPLC's research reports will not know this fact. The consent form included in the appendix (Harwell, D'Amico, Stein & Gatti, 2000) is not a duplicate of the form distributed to teachers, with which I was provided by teachers in two different schools. (Note 3) A cover letter from Deputy Superintendent Bea Johnstone (2000), also omitted from the appendix (Harwell, D'Amico, Stein & Gatti, 2000), an appendix cited in the subsequent study (D'Amico, Harwell, Stein, van den Heuvell, April 2001) refers to this request for teachers' file numbers. The letter states "The survey asks for your name and your teacher folder number, so that the information gained from it can be linked to other data collected in the course of the HPLC study" but that "individual responses will *not* [emphasis in the original] be seen by other members of the District #2 community." Another startling error in the formulation of the questionnaire, also absent from the appendix, is the consent form's listing of Anthony Alvarado, CSD 2's former Superintendent, as a Principal Investigator, with his institutional affiliation given as CSD 2's office ("Consent to act as a participant in a research study" University of Pittsburgh IRB# 980136). Alvarado was not Superintendent when the questionnaire was distributed. Rather, Elaine Fink, a CSD 2 deputy superintendent with whom he continued to collaborate professionally, had replaced him (New York City Board of Education, 2000). It seems apparent that Johnstone's letter to CSD 2 teachers was designed to address teacher apprehensions about the confidentiality of the questionnaire. (Note 4) CSD 2 chapter chairs communicated informally about what they should tell teachers who feared that completed questionnaires would not be confidential (personal communication with two chapter leaders). The poor return rate and the subsequent offer of $500 to any school with a high return rate seem strong evidence that the two chapter chairs who told me that they had informally advised teachers to avoid filling out the questionnaire were not alone.

CSD 2: Not a Typical Urban District

Another key fact about CSD 2 that is not fully addressed in reports by researchers who promote it as a model is the district's access to human and material resources that urban districts typically lack. The "variability" and variation among schools and neighborhoods--the term used to describe CSD 2's demographics (Elmore & Burney, 1999a; Fink & Resnick, 1999)--fails to convey the numerous advantages afforded CSD 2 by the sizable numbers of economically comfortable families who send their children to public school. One study acknowledges that District #2 is "a fairly wealthy urban district... the fourth wealthiest community school district in New York City...and in the upper quartile for urban districts nationally" (Harwell, D'Amico, Stein & Gatti, 2000, p. 9). Surprisingly the implications of this important characteristic are not explored. Reports mention that CSD 2 encompasses a broad swath of Manhattan's wealthiest real estate and most of its prosperous neighborhoods. However, what the reports do not explore is the extent to which the district's concentration of wealthy neighborhoods may spare its central office and many of its schools the psychological and fiscal demands present in most urban school systems, demands that might be summarized as "Keeping students in, gangs out, scores up, alienation down, and the copy machine in working order: Pressures that make urban schools in poverty different" (Metz, 1997).

Research does not explain that CSD 2's showcase, the elementary school touted as having the highest test scores in New York City, is PS 234, in the heart of Tribeca. A ZIP-code by ZIP-code analysis of the New York real estate market found that Tribeca "was the highest priced residential neighborhood in Manhattan last year" (Hevesi, 2002). A report on median sale prices for apartments in Manhattan (Hevesi, 2002) shows how affluent almost all of CSD 2's neighborhoods actually are. Chelsea registered the third highest average price ($1,024,850) for apartments in Manhattan, due to loft conversions that have turned warehouses into art galleries. CSD 2 now contains only one large area inhabited primarily by families living in poverty, Chinatown, populated in great part by newly arrived immigrants. Poverty in the rest of CSD 2 occurs primarily where there are housing projects amid gentrification.

As Tables 1, 2, and 3 show, CSD 2 differs demographically from other New York City districts, especially those with low levels of student achievement. The largest minority population in CSD 2 schools is Asian, and the Asian and white population combined constitute 65% of the students served. In New York City schools, the combined Asian & white figure is 27% (New York City Board of Education, 2000). Data on the school report cards for each school in CSD 2 show that schools in Chinatown serve the highest proportions of students in CSD 2 who are English-language learners, the designation for students who have been in the US for three years or less. (Note 5) In many other districts in New York City, the immigrant population is primarily Spanish-speaking.

New York City designates elementary schools with a "need factor" from 1 to 12, based on the percentage of students categorized as English Language Learners, students identified as eligible for special education, and students eligible for free or reduced lunch. The higher the number, the greater the need of schools for services. As shown in Table 4, more than a third of CSD 2 schools have a "need factor" of only one, two, or three. The "need factor" in these schools indicates that they have a student composition that more closely approximates what teachers would find in the suburban school systems close to the city, where there is no teacher shortage (Institute for Education & Social Policy, 2001). Another important fact that research on CSD 2's professional development fails to address, suggested by scholarship about how teachers' social class influences their work (Metz 1990), is that many teachers attracted to and recently hired by CSD 2 may want to work with administrators, other teachers, and perhaps students, who share their social class origins, aspirations, and world-view. One veteran African-American teacher in a CSD 2 school suggested this possibility to me, noting that minority graduates of "City" (City College, of City University of New York) feel that it is highly unlikely they will be hired by CSD 2 officials. Another factor is that minority graduates of "City" don't want to teach in CSD 2 because they are committed to working in places they see as high need. (Note 6)

Student Achievement in CSD 2

What are the implications of these demographic characteristics of CSD 2? In this section, I examine the extent of racial and economic segregation in CSD 2 schools and compare this to achievement on standardized math tests. I compare data on CSD 2 to equivalent data for the only other school district in all of New York City and New York State to be roughly equivalent in its demographics, District 25 in Queens.

Data on test scores and family income in each school in New York City's public schools, published in *The New York Times* (Goodnough, 2000), reveal the extent to which CSD 2's demographics are *un*representative of other districts in New York City and of urban school districts in New York State as well. Compare, for instance, test results published in October 2000 for CSD 2 in Manhattan and District 8 in the Bronx. Both were reported as enrolling approximately the same number of students, CSD 2 with 2,204, District 8 with 2,374. Yet CSD 2 had 12 elementary schools out of 26 with fewer than 50% of its students qualifying for free lunch, whereas District 8 had one out of 20 (Goodnough, 2000). To pursue the issue of District' 2s representativeness, which I propose as a key consideration in evaluating whether it can indeed be a model for urban school districts elsewhere, I examined demographic data for each county on the New York State Department of Education website (Table 1).(http: //www.emsc.nysed.gov).

Table 1

Comparison of Student Demographics in urban school districts in New York State, Fall 1997*

District	CSD 25, Queens	CSD 2, Manhattan	Yonkers
Total	24,127	22,212	23,968
White	28.2%	31.2%	23.3%
Black	9.1%	13.8%	30.4%
Hispanic	24.4%	21.1%	41.5%
Other	38.4%	33.8%	4.8%
Free/Reduced lunch	48.5%	51.3%	74.7%
Limited English	19.1%	16.9%	16.4%

*Data from the NY State Education Department school district profiles, downloaded on Jan. 31,2002: http://www.emsc.nysed.gov/irts/ch655_99/D662300.html.

As Table 1 shows, the only district in New York State outside of those in New York City that enrolls about the same number of students as CSD 2 is Yonkers. In this comparison two factors that distinguish CSD 2 are its relatively (for an urban district) low percentage of students who qualify for free/reduced lunch and its racial and ethnic mix of students. Examining demographics for each of the community school districts in New York City, I found only *one*, CSD 25 in Queens, that closely resembles CSD 2 in the number of students served (between 20-24,000), the proportion of students reported as eligible to receive free/reduced lunch (between 50-60%), and the student body's ethnic/racial composition (around 10% Black, 30% Hispanic, 30% White, 35% Asian). Note also how the demographics of CSD 2 differ from the characteristics of the New York City school system as a whole: nearly 75% of the children the city school system serves are eligible for free/reduced lunch; more than a third are Black; close to 40% are Hispanic; Asian students (identified as "other" on the school report cards) represent only a little over 10%.

I wish to caution that a complete analysis of the implications of the demographic differences would require close examination of data disaggregated by race and ethnicity that were not publicly available from New York State and New York City until 2002 (email communication, New York State Dept. of Education Data Analyst). The discussion that follows is, therefore, suggestive of the questions that the demographic differences ought to raise. I do not present my discussion as being other than suggestive. Asian students in CSD 2, the largest minority proportionally, would be categorized in John Ogbu's typology (Ogbu, Sept. 1995; Ogbu, Dec. 1995) as "voluntary" (as opposed to "involuntary") minorities.

Ogbu posits that minorities, who emigrate voluntarily, are more likely to experience racism in school and society as barriers to overcome. In contrast, because of their history of oppression by the dominant culture, involuntary minorities are more likely to see racism as a permanent impediment to achievement. Unlike most other districts in New York City and urban districts in New York State, CSD 2's single largest minority population consists of voluntary minorities (Table 2). Both Ogbu's typology and the categories used to report demographics in New York schools (Black, White, Hispanic, Other- Asian and Pacific Islander) obscure very important differences among these populations. As Cooper and Denner (1998) noted, a limitation in the analytical framework Ogbu employs is its lack of emphasis on variation and change within communities, especially upwardly mobile ethnic minority families and children. Gibson (1997) argued that Ogbu's typology fails to account for intragroup variability and is too "dichotomous, too deterministic, and in danger of contributing to stereotypical images..."(p. 322). She faults the theory for not taking into account generational and gender differences, as well as lacking explanatory power to account for the experience of groups such as Mexican Americans, who share elements of both categories. Moreover, she contends, the theory fails to take into account school effects and human agency. While acknowledging the limitations in Ogbu's typology, I suggest that his theory helps illuminate why school and instructional practices that are successful with one group of students may not be equally effective with another, and thus his typology is germane to discussions of CSD 2's achievements.

Table 2

Comparison of Demographics of CSD 2 and CSD 25

	NYC	CSD 25, Queens	CSD 2, Manhattan
Total enrollment	NA on report card	24,499	21,559
White	15.3%	26.3%	31.9%
Black	34.2%	8.4%	13.8%
Hispanic	38.9%	25.3%	20.1%
Other -Asian, Pacific Islanders, Alaskan Natives, Native Americans	11.5%	40.1%	34.2%
Free/Reduced lunch	74.1	50.4%	59.9
Recent immigrants	7.3%	12.3%	9.2%
Limited English	NA on report card	4557 students	2940 students

Another, thornier problem with applying Ogbu's analysis is that New York City Board of Education's demographic data do not parallel Ogbu's categories. For example, the category "English Language Learners" describes recent immigrants both from Puerto Rico and Peru; "Hispanic" students are those with Spanish surnames; students from British Guayana, who are "voluntary" minorities might be labeled "Black," placing them with African Americans, "involuntary" minorities. Although I group students categorized as "Hispanic" in with those categorized as "Black" in the analysis, I caution that some of these students may share the cultural framework of reference of "voluntary" and not "involuntary minorities."

I divided CSD 2's 25 elementary schools into two categories: those serving a population more than 50% combined Black and Hispanic students with a "need" factor of 7 or greater, and those with fewer than 50% Black and Hispanic students and a "need" factor less than 7. In CSD 2 there are 11 schools with a "need" factor over 7. Five of these 11 schools are high-poverty schools in Chinatown. Table 3 shows the breakdown of schools in both districts, according to these criteria. I have eliminated the schools in Chinatown from this comparison to look at achievement of majority Black and Hispanic schools in both districts.

Elmore and Burney (1999b) note that 18 schools in CSD 2 have populations more than two-thirds African-American, Hispanic, and Asian, while four have populations that are more than two-thirds white . As noted earlier, the implications of this finding of "variability" are not explored further, in particular the extent to which achievement in CSD 2 schools correlates with their racial and social stratification. As is evident from Table 3, only one elementary school, PS 11, has a student population that mirrors the district's demographics. In CSD 2, 5 of the 11 schools with a "need" factor 7 and above are in Chinatown and have a population that is more than 70% Asian. For instance PS 42, with a "need" factor of 10 (94.2% of its students receive free/reduced lunch and 18.9% are ELL) enrolls 88% Asian students. What is not evident in statistical analyses is that PS 11 also has a very large "gifted and talented" program in which almost all of its White students are enrolled (personal communication, CSD 2 administrator). Hence, the one elementary school that is demographically representative of the district's enrollment houses two different schools, one serving White students in its "gifted and talented" students, the other Black and Hispanic students. The school's scores are reported in the aggregate.

Table 3

Elementary Schools Serving a Population More Than 50% Combined Black and Hispanic Students with a "Need" Factor of 7 or Greater

School	"Need"	% Hispanic	% Black
CSD 25: PS 201	10	29%	40%
School	"Need"	% Hispanic	% Black
CSD 2: PS 11	7	31%	26%
CSD 2: PS 33	10	52	27
CSD 2: PS 51	10	61	18
CSD 2: PS 111	9	66	16
CSD 2: PS 126	7	43	30
CSD 2: PS 151	10	43	30
CSD 2: PS 198	8	52	26

Note: Data from NYCBOE website, http://www.nycenet.edu/daa/Mrr/districts.

What is consistently referred to as "variability" or "variation" in school demographics in CSD 2 is actually a euphemism for a familiar phenomenon in US schools: racial segregation (Orfield & Eaton, 2003). The high degree of racial and social stratification is especially noteworthy in light of comparison to CSD 25. With approximately the same demographics as CSD 2, CSD 25 has only *one* school that is as racially segregated as *eleven* schools in CSD 2.

"Student Need" is a ranking of elementary schools into one of 12 categories based on the percent of students eligible for free lunch, percent of tested students who are in full-time and part-time special education programs, and the percent of tested students who are English Language Learners (ELL). The higher the numbers of students in these categories, the higher the school's "need" factor. To understand the significance of "need," two schools (in CSD 25) designated as having "need" factors of "3" and "10" are compared in Table 4. Most schools in CSD 25 have a "need" factor within the range of 3-7 (17 of 23 schools fall within this range; 6 schools are outliers). CSD 2 has only 9 of 24 that fall within this range. Seventeen of its schools fall outside this range. The comparison indicates that CSD 2's schools are far more stratified than those in CSD 25, a district with a student enrollment that is equivalent in terms of the demographic categories used by the state.

Table 4

Examples from CSD 25 of "Student Need" Categories 3 and 10 for Two Elementary Schools

School's need factor	% students receiving free/reduced lunch	% special ed.	% ELL
CSD 25 School A "Need" of "3"	40%	4%	8%
CSD 25 School B "Need" of "10"	73%	36%	5%

Comparing achievement between schools in both districts, I used scores on the New York State fourth grade math test in 2000, reported in the *New York Times* (Goodnough, 2000). I compared scores of only those schools serving a majority of Black and Hispanic students. In the New York State tests, scores of level 1 and 2 indicate that the student is "not meeting standards." The results of this comparison are shown in Table 5.

Table 5

Scores on the NYS Fourth-Grade Math Test Only in 2000
(Elementary schools serving a population more than 50% combined Black and Hispanic students with a "need" factor of 7 or greater)

School	perf level 1	perf level 2	perf levels 3 & 4	"Need" factor	% Hispanic	% Black
CSD 25: PS 201	17%	41%	43%	10	29	40
School	**perf level 1**	**perf level 2**	**perf levels 3 & 4**	**"Need" factor**	**% Hispanic**	**% Black**
CSD 2: PS 11	16%	16%	68%	7	31	26
CSD 2: PS 33	26%	41%	33%	10	52	27
CSD 2: PS 51	19%	59%	27%	10	61	18
CSD 2: PS 111	26%	47%	27%	9	66	16
CSD 2: PS 126	9%	45%	45%	7	39	20
CSD 2: PS 151	15%	58%	28%	10	43	30
CSD 2: PS 198	4%	20%	76%	8	52	26

Source: Reported in the *NY Times*, 15 Oct. 2000, (City section, pp. 14-16)

In Table 6, I continue the comparison of school-wide test scores, using data from the school report cards published on the New York City Board of Education website and including the number of students tested. These test scores, unlike the others I have analyzed, are for math scores in grades 3-8. Two of the schools in CSD 2 are K-8 schools, PS/IS 33 and PS/IS 111. With the exception of PS 11, which has a large "gifted and talented program" of mainly white students, *only one school in CSD 2 has a significantly higher proportion of students "meeting standards" in math than in the school with similar demographics in CSD 25.* Indeed, several CSD 2 schools *do not perform as well* as PS 201 in CSD 25.

Table 6

Percentages of Students "Meeting the Standard" (i.e., Perf. Levels 3 and 4) on City-wide Math Tests in Grades 3,5,6,7 and State Math Tests in Grades 4 and 8 (including students taking the test in translation)

School	Meeting standard	"Need" factor	% Hispanic	% Black
CSD 25: PS 201	31.5% (235 tested)	10	29%	40%
School	**Meeting standard**	**"Need" factor**	**% Hispanic**	**% Black**
CSD 2: PS 11 (34% white; 8% Other)	50.6% (233 tested)	7	31%	26%
CSD 2: PS/IS 33 (White 4%; Asian 37%)	17.3% (191 tested)	10	52	27
CSD 2: PS 51	22.5% (145 tested)	10	61	18
CSD 2: PS/IS 111	22.5% (516 tested)	9	66	16
CSD 2: PS 126	32.6% (285 tested)	7	43	30
CSD 2: PS 151	25% (115 tested)	10	43	30
CSD 2: PS 198	53.4% (131 tested)	8	52	26

Several questions are posed by this comparison with CSD 25, one of the most critical being what data disaggregated by race and ethnicity reveal about achievement. With the exception of the studies based on the questionnaire requiring teachers to self-identify with file numbers (Harwell, D'Amico, Stein & Gatti, 2000), reports by researchers promoting CSD 2 as exemplary do not address this question. (Note 7) But in the one study that does attempt to use data disaggregated by race and ethnicity, the investigators state their primary research question as this: "Are teachers with strong professional development participation patterns more likely to have closed achievement gaps?"(Harwell, D'Amico, Stein & Gatti, 2000, p. 19). The answer: "In summary, engagement in professional development, as measured by this questionnaire and reported by the 62 respondents, does not appear to have significant influence on student achievement in either literacy or mathematics" (Harwell, D'Amico, Stein & Gatti, 2000, p. 22).

Conclusions

In spite of statements by researchers looking for evidence that CSD 2's policies have indeed boosted achievement, there are no data to support such claims (Harwell, D'Amico, Stein & Gatti, 2000). Indeed, no data support even the more modest claim of the "generally positive picture" of systemic reform in CSD 2 (Elmore & Burney, 1999a, p. 3). How then could researchers promote CSD 2's instructional development practices as unusually successful or its investment in staff development tied to national standards as a model to be emulated?

Research on CSD 2 exemplifies the problems that arise when researchers fail to maintain the independence and critique that Bourdieu (1998) demanded of intellectuals. The inter-dependence of district leaders and researchers, combined with the exclusion of dissenting perspectives, obscured key questions about CSD 2 practices that need to be explored before they can or should be replicated. From the formulation of research design, to data collection, to presentation of findings, research on CSD 2 appears to have shown a marked disregard for alternative perspectives and local knowledge. As a result, insights that might have contributed to district officials' and researchers' learning have been ignored . Researchers have published reports that have reinforced the belief among CSD 2 officials that their work is a model for the entire New York City system because it is "leading New York City in implementing Standards" (Harwayne, 1999). But comparison of existing data for CSD 25, knowledge of the social-contextual factors such as CSD 2's access to human and material resources other districts in the city lack, and the inattention to disaggregation of individual achievement according to race and ethnicity, indicate that the representation of CSD 2's practices as exemplary by researchers is unsubstantiated. The "labor-management" strategy that resulted from close relationships and consensus between high level union officials and district administrators caused significant turmoil that was not reported in the research. The consensus may not be replicable elsewhere, indeed, was probably disrupted in CSD 2 with the election of a new "district rep." The rise in achievement levels since 1989 may be due to changes in the district's demographics and not to a focus on instruction or professional development linked to national standards. Research that has attempted to link achievement to professional development has failed to find evidence of correlation, let alone establish causation.

It may be that CSD 25 is, in fact, just as promising a model of school improvement as CSD 2's. Its elementary schools are far less segregated and stratified by income than are CSD 2's. Its test scores are equivalent to those in CSD 2. It is interesting to note that the statement of its Superintendent in the 1999-2000 suggests a stance towards students, parents, and community sharply different from that promulgated by CSD 2:

> ... We teach "children not merely subjects." To support this goal, the District and the Community School Board work closely to provide an integrated, holistic, comprehensive educational program which motivates and engages all students, and provides the optimum opportunity for every child to achieve state and city academic standards... . Staff are supported by professional development activities designed to help them hone their instructional skills. Parents and community members are actively involved in all schools and are recognized as valuable resources.

It may be that CSD 2 has pioneered practices that should be replicated, as the researchers who have promoted it have concluded. However, the opposite conclusion is equally plausible. It was the task of research to explore both possibilities, but the role of cheerleader seems to have over-ridden the demands of scholarship.

Notes

1. As of 2003, the NYC Board of Education has been renamed the "Department of Education" and control over the city schools given to Mayor Bloomberg. He and his appointed Chancellor have submitted a proposal to the state legislature to merge the 32 community school districts into 10. CSD 2's present superintendent has been selected to head one of the 10 new districts.

2. Information about the group publishing the report on the math curriculum is available from Elizabeth Carson, ecarson@nyc.rr.com.

3. I informed Richard Elmore and Lauren Resnick about the concerns raised to me by CSD 2 teachers about

the use of file numbers and Alvarado's presence as a PI. Lauren Resnick responded (letter, 17 July 2000), and Stein and Resnick met with me in April 2001. According to notes I took after our meeting, the objections I raised to the conduct of research, namely that it deepened a climate of fear, were dismissed. Resnick acknowledged the possibility of "bad design." Stein noted that their findings actually contradicted claims being made about CSD 2's success. However both Stein and Resnick rejected my proposal to clarify publicly that HPLC's latest research told a different story about CSD 2 from the one that had been widely publicized in earlier work. Resnick explained that their role had ended with the study's completion.

4. Another item omitted from the Appendix and not mentioned in the reports was a notice distributed to teachers, signed "The HPLC Research Team" with the HPLC address, phone, fax, and website. It announced a reward of $500 to schools in which 90% of the teachers returned questionnaires. The notice also informs teachers that the extra consent form to serve as their personal copy, included in the original packet with Johnstone's cover letter, "is the wrong version" and should be discarded.

5. I examined the 1999-2000 report cards for all elementary schools in CSD 2. They and the District report cards are available at http://www.nycenet.edu/daa.

6. Questionnaires returned to researchers show a preponderance of White, female, middle-class respondents (D'Amico, Harwell, Stein, Van den Heuvel, April 2001). Curiously, researchers on CSD 2 have never investigated the extent to which district hiring screens out teachers and principals whose professional beliefs differ from those of the district leadership, and the ways those beliefs correlate with social class or race.

7. The investigators secured achievement and demographic data for individual students in District #2 through the Division of Assessment and Accountability of the Office of the Deputy Chancellor of Instruction of the New York City Board of Education. Disaggregated data have heretofore been made available only to researchers working with the Division of Assessment and Accountability, as I learned when I attempted to secure it for this study (email message from data analyst at DAA).

References

American Educator. (1999-2000, Winter). It's all about teaching and learning. *American Educator*, pp. 14-19.

Bourdieu, P. (1998). *Acts of resistance. Against the tyranny of the market.* New York: The New Press.

Cooper, C. R., & Denner, J. (1998, April). Theories linking culture and psychology: Universal and community-specific processes. American Educational Research Association. San Diego.

Cuban, L. (1988). Researchers advising policymakers: A word to the wise. *Educational Psychologist, 23*(3), 287-293.

D'Amico, L., Harwell, M., Stein, M. K., & van den Heuvel, J. (2001, April). Examining the implementation and effectiveness of a district-wide instructional improvement effort. American Educational Research Association. Seattle.

D'Amico, L., van den Heuvel, J., & Harwell, M. (2000). *Multiple Perspectives on Instructional Improvement in Literacy and Mathematics.* High Performance Learning Communities Project, contract #RC-96-137002. Learning Research and Development Center, University of Pittsburg.

Elmore, R. F. (1999-2000, Winter). Building a new structure for school leadership. *American Educator*, pp. 6-13.

Elmore, R. F., & Burney, D. (1997a). *Investing in teacher learning: Staff development and instructional improvement in Community School District #2, New York City* [ERIC Document Reproduction Service, ED416203]. New York: National Commission on Teaching & America's Future. Teachers College, Columbia University (62 pages).

Elmore, R. F., & Burney, D. (1997b). *School variation and systemic instructional improvement in Community School CSD 2* [Http://www.lrdc.pitt.edu/hplc/hplc.html]. Pittsburgh: University of Pittsburgh, High Performance Learning Communities Project, Learning Research and Development Center.

Elmore, R. F., & Burney, D. (1999). *School variation and systemic instructional improvement in Community School District #2, New York City* [The CEIC Review. Volume 8, Number 1]. Philadelphia: National Center on Education in the Inner Cities. Temple University Center for Research in Human Development and Education.

Elmore, R. F., & Burney, D. (1999, February). School variation and systemic instructional improvement in Community School District #2, New York City. In A. Grant & M. Calabazana (Eds.), *The CEIC Review*. The CEIC Review. Philadelphia: The National Center on Education in the Inner Cities. Temple University Center for

Research in Human Development and Education.

Fink, E., & Resnick, L. B. (1999). Developing principals as instructional leaders. *HPLC reports* [Online]. Available: www.lrdc.pitt.edu/hplc/hplc.html (Accessed 12 November 2001).

Freedman, S. (1987). Teacher 'burnout' and institutional stress. In J. Ozga (Ed.), *Schoolwork. Approaches to the labour process of teaching* (pp. 133_145). Philadelphia: Open Univ. Press. Milton Keynes.

Knapp, M. (1995). The teaching challenge in high_poverty classrooms. In M. Knapp & & associates (Eds.), *Teaching for Meaning in High_Poverty Classrooms* (pp. 160_182). New York: Teachers College Press.

Gibson, M. (1997, September). Exploring and explaining the variability: Cross-national perspectives on the social performance of minority students. *Anthropology and education, 28*(3), 318-329.

Goodnough, A. (2000, 15 October). How grades four and eight fared on tests. *The New York Times*, pp. CY 14-16.

Harwell, M., D'Amico, L., Stein, M. K., & Gatti, G. (2000). Research contract #RC-96-137002 with OERI. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.

Hevesi, D. (2002, 17 May). TriBeCa is priciest neighborhood. *The New York Times*, p. B6.

Institute for Education & Social Policy. (2001). *Distributing teacher quality equitably: The case of New York City*. New York: Institute for Education & Social Policy, New York University.

Johnstone, B. (2000, 1 May). [Letter to District #2 teachers distributed by principals]. New York City.

Metz, M. H. (1990). How social class differences shape teachers' work. In M. W. McLaughlin, J. E. Talbert & N. Bascia (Eds.), *The contexts of teaching in secondary schools. Teachers Realities* (pp. 40-107). New York: Teachers College Press.

Metz, M. H. (1997, March). "Keeping Students in, Gangs Out, Scores up, Alienation Down, and the Copy Machine in Working Order: Pressures That Make Urban Schools in Poverty Different." American Educational Research Association. Chicago.

New York City Board of Education. (2000). *Second Annual Report: Evaluation of the Performance Driven Budgeting Initiative*. New York City: Author.

Ogbu, J. U. (1995, September). Cultural problems in minority education: Their interpretations and consequences-Part One: Theoretical background. *The Urban Review, 27*(3), 189-205.

Ogbu, J. U. (1995, December). Cultural problems in minority education: Their interpretations and consequences-Part two: Case studies. *The Urban Review, 27*(4), 271-297.

Orfield, G. & Eaton, S. E. (accessed 11 March 2003). Back to segregation . http://www.civilrightsproject.harvard.edu/research/articles/reseg.php

Resnick, L., & Harwell, M. (1998). *High performance learning communities: CSD 2 achievement* [OERI research contract #RC-96-137002 deliverable]. Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Resnick, L. B., & Harwell, M. (2000). *Instructional variation and student achievement in a standards-based education district.* Pittsburgh, PA: Learning Research and Development Center.

Stein, M. K., D'Amico, L., & Johnstone, B. (1999, April). District as professional educator: Learning from and in New York City's Community School District #2. American Educational Research Association. Montreal.

Weiner, L. (2002, May). Assessing systemic's reform "learning by all": Who evaluates learning by policy analysts? *Educational Policy, 16* (2): 239-263.

Weiner, L. (1998, Summer). Albert Shanker's legacy. *Contemporary Education*, *69* (4), 196-202.

Zuckerman, J. B. (2001). Queering the life of a progressive, urban, elementary school: Genealogical ghost stories [Dissertation], New York: Teachers College, Columbia University.

About the Author

Lois Weiner
Lois Weiner is a Professor of Elementary and Secondary Education at New Jersey City University, 2039 Kennedy Blvd., Jersey City, NJ 07305. Email correspondence should be addressed to LWEINER@NJCU.EDU.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
University of Washington
Seattle Pacific University

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)

CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 1160 times since August 7, 2003

Volume 11 Number 28 **August 7, 2003** **ISSN 1068-2341**

Reforms, Research and Variability: A Reply to Lois Weiner

Lauren B. Resnick
University of Pittsburgh

Citation: Resnick, L. B. (2003, August 7). Reforms, research and variability: A reply to Lois Weiner. *Education Policy Analysis Archives, 11* (28). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n28/.

Abstract
Lois Weiner (2003) argues that the research reports from High Performance Learning Communities (HPLC) were biased because of the close working relationships between the researchers and the leaders of the Community School District Two (CSD2) reform. Contrary to any claims otherwise, this relationship was quite open and acknowledged. The intent of the HPLC investigation was always to link scholars and practitioners in a new form of research and development in which scholars became problem-solving partners with practitioners. There are important issues about how to profitably conduct such "problem-solving" research. These issues are worth substantial attention from the communities of researchers and practitioners as collaborative research/practice partnerships proliferate. Serious studies of such partnerships are needed, going well beyond the anecdotal attacks offered by Weiner in her article.

Dr. Weiner's (2003) article is at once an analysis of data on demographics and achievement in Community School District Two (CSD2) in New York City and an attack on the research strategy (and by implication the research ethics) of the High Performance Learning Communities (HPLC) project that I co-directed, along with Richard Elmore and Anthony Alvarado. Her paper begins with what can only be construed as a personal attack on the researchers and practitioners of the HPLC project. The attack is inherent in the title of the paper, in the way quotes are used and in the personal story of Weiner's own interest that threads through the introduction but is never fully documented. Nevertheless, I welcome Dr. Weiner's effort to provide new data and a fresh perspective on work to which we devoted substantial professional effort during the period 1996-2001.

Dr. Weiner points out that CSD2 was not a typical urban district in terms of its demography. She is absolutely right about this. As we have noted in most reports on HPLC research, the district sits in the midst of some of the greatest concentrations of wealth in the nation, and a noticeable (although minority) portion of middle class parents send their children to CSD2 schools. At the same time, the district has large numbers of students of color and families of poverty (as measured by eligibility for free and reduced lunch), as well as immigrant students who are in the process of learning English.

Dr. Weiner's reports of the demographics of the district as a whole and schools within the district roughly match the data we have collected and reported in several papers over the course of our five-year study. Perhaps most important to her argument is that schools in CSD2 did not uniformly represent the demographics of the district as a whole. That is correct. CSD2 contained "rich" schools and "poor" schools, schools with very few children of color and others almost entirely filled with minority students. The variability among schools was never in question. What is important to ask is whether CSD's unique (at the time) system of curriculum leadership and professional development within schools led to learning gains—especially in "high need" schools.

Dr. Weiner addresses this question by comparing CSD2 schools with relatively high need ratings (7 schools in all) with a *single* high need school in District 25 in a single year on a single test in one subject matter. Overall, the CSD2 schools did not outperform the District 25 school. Dr. Weiner seems to imply that we should therefore conclude that CSD2's program of curriculum and professional development was not effective. Possibly—but it pays to look at more evidence than she provides. HPLC conducted a number of analyses of CSD2 academic performance, both for the district as a whole and school-by-school. Summaries of most of these analyses appear in the 2001 *Final Report* of the project.(Note 1)

One HPLC analysis examined changes over time in reading and mathematics during the period 1992 to 1998—a period in which the CSD2 curriculum and professional program was being put into place and expanded, and during which a stable test in each subject was being used in New York City. In 1993—the first year of New York's renormed math achievement test, just under 70% of CSD2 students were at or above grade level in math; in 1998 , about 82% of CSD2 students were at or above grade level. The story is similar for reading: Scores rose from just under 60% at or above grade level in 1992 to about 72% in 1998.

This gradual rise in overall achievement could have resulted from a change in overall district demographics resulting from more middle class students attending CSD schools. But it did not. According to our data, during this period the percentage of students in the district eligible for free or reduced lunch remained stable at about 53%.

The overall improvement also might have resulted mainly from nudging students already near "grade level" over the mark into performance level 3, leaving the students in greatest need behind. To check this, we analyzed achievement quartile-by-quartile wherever such data was available in several successive years on the same test. In the period between 1996 and 2000, the proportion of CSD2 students testing in the bottom quartile in reading fell a bit every year—to a low of just over10% by 2000. Math drops were smaller, perhaps because the CSD2 math curriculum and professional development system was introduced later and might not yet have fully taken root.

Unfortunately, we were not able to obtain detailed enough data on other districts in New York to make comparisons with them. But we were able to use the variability among schools within CSD2 to examine whether the leadership's curriculum and professional development system can be credited with raising achievement, especially for children with the greatest academic need. Using questionnaires and ratings of classroom instructional quality to assess the extent of engagement in the CSD2 program, our studies showed that deep teacher engagement in professional development and faithful implementation of the district's literacy and math programs both raised overall achievement and reduced the connection between achievement and socioeconomic status. It is interesting that a similar finding for mathematics, using a different curriculum but a similar professional development system, was reported for the Pittsburgh Public Schools during the period 1996-98. (Note 2)

We are still left with the question of whether there was something special about the mix of students and schools in CSD2 that might have made it easier than elsewhere to effect the kinds of learning changes that the leadership sought. Here Dr. Weiner makes an important contribution in calling to attention the fact that there are important cultural differences among minority and English learning groups. As she points out, several of the high poverty schools in CSD2 were Chinatown schools. She suggests that Chinese immigrants are, in Ogbu's terminology, "voluntary" immigrants, and their children perhaps more likely to participate actively in the opportunities offered by schools. This and other possible cultural difference between the Chinatown schools and other high poverty and minority schools certainly warrant further investigation. Meanwhile, however, Weiner's data make it clear that Asian students and schools were not the only minorities who did well in the CSD2 reform effort. Note, for example that the school with the highest overall academic performance in Dr. Weiner's Table 5 (PS 198 in CSD2), had a population of 52% Hispanic and 26% Black students. Thus, interesting as the "Asian question" is, does not call the overall record of CSD2 into question.

This brings me back to what appears to be Dr. Weiner's main point: That the research reports from HPLC were biased because of the close working relationships between the investigators and the leaders of the CSD2 reform. There was never a secret about this relationship. Indeed the intent of the HPLC investigation from the start was to link scholars and practitioners in a (then) new form of research and development in which scholars became problem-solving partners with practitioners and practitioners accepted the responsibility of collecting evidence in as unbiased a manner as possible, using it to refine and—when necessary—alter their theories of action.(Note 3) Our goal was to deeply document, analyze and understand the actual practices of the CSD2 reform. We conducted extensive interview and observation studies of professional development and classroom practice as well as the studies discussed here that examined impact on student learning. Variability—among students, teachers and schools—was a central object of investigation and analysis throughout.

HPLC was not a simple undertaking and there were difficulties encountered along the way. Dr. Weiner points to one of them—an unwillingness of many teachers to return a questionnaire when they felt that anonymity could not be guaranteed because of the relationship between the researchers and the district leaders. Problems of that kind are easy to recognize and to address. There are in addition deeper issues about how to profitably conduct such "problem-solving" research, however—for example, the subtle ways in which question are formulated, or avoided, because of the common perspectives that emerge in long collaborations. These issues are worth substantial attention from the research and practice communities as collaborative research/practice partnerships proliferate. Serious studies of such partnerships are needed, going well beyond the anecdotal attacks offered by Dr. Weiner in the opening sections of her paper.

At the same time, it is critically important to conduct more "arm's length" research on reform programs or other interventions that appear to be succeeding. The HPLC project did not claim to be an arm's length investigation, but all parties to it would welcome such investigations. CSD2 as such no longer exists (having been absorbed into a much larger instructional Region in New York's recent school reorganization). But many of the ideas pioneered in CSD2 are now being tried in districts across the country. There is, thus, plenty of opportunity both for collaborative problem-solving research of the HPLC variety and more arm's length evaluations. Such studies will tell us what aspects of the CSD2 effort can "travel" well to other environments, what effects they have on various populations of students and educators, and—of utmost importance—what revised or totally new theories of action are likely to meet the demands for increased academic achievement across a broad spectrum of the school population.

Notes

1. *High Performance Learning Communities Project: Final Report* (September 15, 2001). Learning Research and Development Center, University of Pittsburgh, September 15, 2001. (Available on the HPLC website: http://www.lrdc.pitt.edu/hplc)

2. Briars, D., & Resnick, L. B.. *Standards, assessments – and what else? The essential elements of standards-based school improvement.* CRESST Technical Report, 2000.

3. This form of *Problem Solving Research and Development* was recommended in a National Academy of Education report to OERI as one important means of bringing research and practice into closer interaction See Brown, A.L. & Greeno, J.G. (Eds.) *Recommendations regarding research priorities: An advisory report to the National Educational Research Policy and Priorities Board.* National Academy of Education. Spring, 1999.

Reference

Weiner, L. (2003, August 7). Research or "cheerleading"? Scholarship on Community School District 2, New York City. *Education Policy Analysis Archives, 11*(27). Retrieved [date] from http://epaa.asu.edu/epaa/v11n27/.

About the Author

Lauren B. Resnick
Director and Senior Scientist
Learning Research and Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Phone: 412-624-7020
Fax: 412-624-9149
Email: resnick+@pitt.edu

Lauren B. Resnick is an internationally known scholar in the cognitive science of learning and instruction. Her recent research has focused on school reform, assessment, effort-based education, the nature and development of thinking abilities, and the relation between school learning and everyday competence. Her current work lies at the intersection of cognitive science and policy for education. Dr. Resnick founded and directs the Institute for Learning, which focuses on professional development based on cognitive learning principles and effort-oriented education. She is co-founder and co-director of the New Standards Project, which has developed standards and assessments that have widely influenced state and school district practice. Resnick was a member of the Commission on the Skills of the American Workforce and served as chair of the assessment committee of the SCANS Commission and of the Resource Group on Student Achievement of the National Education Goals Panel. She has served on the Commission on Behavioral and Social Sciences and Education and on the Mathematical Sciences Education Board at the National Research Council. Her National Academy of Sciences monograph, *Education and Learning to Think*, has been influential in school reform efforts, and her widely circulated Presidential Address to the American Educational Research Association, "Learning In School and Out," has shaped thinking about youth apprenticeship and school-to-work transition. Dr. Resnick is Professor of Psychology at the University of Pittsburgh, where she directs the Learning Research and Development Center. Educated at Radcliffe and Harvard, she received the 1998 E. L. Thorndike Award from the American Psychological Association and the 1999 Oeuvre Award from the European Association for Research on Learning and Instruction.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University

Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz Garcia (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil
(AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 584 times since August 19, 2003

Volume 11 Number 29 **August 19, 2003** **ISSN 1068-2341**

Comments on Weiner, Resnick and Scientific Debate

Jonathan Goodman
New York University

Citation: Goodman, J. (2003, August 19). Comments on Weiner, Resnick and scientific debate. *Education Policy Analysis Archives, 11* (29). Retrieved [date] from http://epaa.asu.edu/epaa/v11n29/.

Abstract
Lois Weiner (2003) and Lauren Resnick (2003) have advanced substantially different views of the success of the reforms undertaken by Community School District Two (CSD2) in New York city. Weiner's position *vis a vis* District Two has probably conferred a greater measure of objectivity to her views. Criticisms of scholarly work, even when sharply worded, are neither personal nor unscientific; indeed they are quite common in all the sciences.

Lois Weiner, Lauren Resnick, and the *Archives* are to be commended for initiating a public peer-reviewed debate on the studies used to support what is unfortunately called "reform" education, fuzzy math and whole language reading. As is becoming clear nationally, statistics about educational programs are not always what they seem. The scientific method includes the axiom that adversarial scrutiny of data is the only path to correct conclusions. Researchers challenging fuzzy math have often been denied access to education journals and their criticisms have gone unanswered and unrecognized in the education community (See the web site http://mathematicallycorrect.com for instances of this.)

Another axiom of scientific objectivity is that the most reliable evidence for a particular reform is not likely to be from the proponents of that reform. In medical practice, for example, a new procedure must be tested in clinical trials by researchers other than those who propose it before it is accepted. Precisely because Weiner was not involved in implementing the District 2 reforms, her analysis of the results is less likely to be biased, though it might be biased for other reasons.

Anyone who actually lived in District 2 and had children in District 2 schools during the period in question—as I did—would be aware of the huge shift in demographics that Weiner pointed out. For one thing, at least two new schools serving well the "well-to-do" (PS 234 and PS 98/IS 89) began operations. Moreover, a number of magnet schools (Lab, Salk, School of the Future, etc.) began drawing top students from throughout the city. It would have been more informative to present results from individual schools whose student populations were more stable. Having seen such data but not having access to them now, I do not recall that it was nearly as positive as the overall numbers quoted by Resnick.

It is hard to agree with Resnick's complaints about the tone of Weiner's article. Criticism of the structure of a scientific study is not a personal attack on the ethics of the researchers. Scientific journals that engage in debate, such as *Physical Review Letters*, regularly use such language. The fact that scientific opinions are colored by personal factors does not make them wrong. Having interacted with officials of District 2 over the years, I can testify that it is difficult to describe their actions dispassionately. Weiner has done remarkably well, given her conclusions.

Reference

Resnick, L. B. (2003, August 7). Reforms, research and variability: A reply to Lois Weiner. *Education Policy Analysis Archives, 11*(28). Retrieved August 14, 2003, from http://epaa.asu.edu/epaa/v11n28/.

Weiner, L. (2003, August 7). Research or "cheerleading"? Scholarship on Community School District 2, New York City. *Education Policy Analysis Archives, 11*(27). Retrieved August 14, 2003, from http://epaa.asu.edu/epaa/v11n27/.

About the Author

Jonathan Goodman is Professor of Mathematics at the Courant Institute of New York University.

Email: goodman@cims.nyu.edu
Homepage: http://www.math.nyu.edu/faculty/goodman

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
California State Univeristy–Los Angeles

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 677 times since August 25, 2003

Volume 11 Number 30 **August 25, 2003** **ISSN 1068-2341**

Embracing Pedagogical Pluralism:
An Educator's Case for (at Least Public) School Choice

David J. Ferrero
Bill & Melinda Gates Foundation
Seattle, Washington (U.S.)

Citation: Ferrero, D. J. (2003, August 25). Embracing Pedagogical Pluralism: An Educator's Case for (at Least Public) School Choice *Education Policy Analysis Archives, 11*(30). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n30/.

Abstract

Pedagogical and curricular beliefs and commitments are expressions of deeper philosophical and ideological worldviews that empirical research can sometimes modify but not ultimately eliminate. The pluralism these views produce is *reasonable* in that they all represent plausible interpretations of liberal-republican values and professional standards of practice; they should be granted some room to flourish under a system of carefully regulated autonomy and choice. Three objections to a conception of school choice grounded in a notion of reasonable pluralism among educational doctrines are addressed: 1) that it would undermine educators' efforts to secure status for themselves as professionals by admitting that "best practices" in education offer rough guidance at best; 2) that it would leave parents and students vulnerable to quackery; 3) that it abandons the common school tradition and its aspirations. I conclude with an examination of why the conceptual basis on which a society designs a system of choice makes a difference.

Disagreement and debate within a professional community can be healthy. It sharpens thinking, stimulates inquiry, and expands knowledge. A community without such stimulative controversy would be moribund. At the same time, too much disagreement about too many fundamentals leads to schism. Education is characterized by both kinds of controversy. Competing hypotheses about how children learn have spurred wide-ranging research that has converged on some core principles to guide professional socialization and practice, at least roughly. On the other hand, the broad consensus about purposes and processes has done little to abate the internecine battles over what these principles imply for practice. The last two decades alone have given us the "reading wars," the "math wars," and the "culture wars," as well as fierce battles over standards, what constitutes a "qualified teacher," and more broadly, continued skirmishes in the Hundred Years' War between various types of educational "traditionalists" and "progressives." These battles have been engaged by educators and non-educators alike, and a staggering amount of time and energy have gone into waging them. I don't know how to measure the impact of these perennial conflicts on children, communities, and the teaching profession itself. But I do know that civil wars are never healthy, and have wondered for a long time now what might be accomplished if these passions and energies were channeled in more productive directions.

The differences that set educator against educator are intractable, and all the research in the world will not settle their disputes. This is because the most important questions that divide them are normative rather than empirical. These divisions go deep. Pedagogical and curricular beliefs are extensions of more comprehensive philosophical doctrines that are in turn colored by ideological ones. In other words, educational doctrines reflect metaphysical, epistemological, and ethical commitments conditioned in part by identity and a certain understanding of history and society. They constitute what John Rawls (1993) has termed "comprehensive moral doctrines," that is "conceptions of what is of value in human life, as well as ideals of personal virtue and character, that are to inform our . . . conduct (in the limit of our life as a whole)" (p.175). As such, differing pedagogical belief systems ought to receive the same treatment as other forms of pluralism under liberal-democratic regimes—that is, tolerance within reasonable bounds. And given the depth of educators' commitments to competing and mutually incommensurable conceptions of their vocation, it would seem that educators have much to gain in terms of satisfaction and effectiveness from an arrangement that gave them greater freedom to create schools according to their ideals with like-minded colleagues—perhaps enhancing, rather than diminishing, their status as professionals. Such an arrangement would entail a degree of autonomy for educators to assemble for purposes of creating schools that realize their ideals within broadly established political and pedagogical limits. It correspondingly entails *choice* on the part of both educators and families, because the different kinds of schools created under such an arrangement would correspond to the values and needs of different students and parents. In short, a system of school choice based on differing conceptions of good schooling would be a good thing for educators—as well as for families and communities—because it could foster the creation of more cohesive learning communities built on common beliefs about teaching and learning.

The argument proceeds in three steps. First I review, in a schematic and oversimplified way, how pedagogical and curricular beliefs and commitments are expressions of deeper philosophical and ideological worldviews that empirical research can sometimes modify but not ultimately eliminate. I then argue that the pluralism these views produce is nonetheless *reasonable* in that they all represent plausible interpretations of liberal-republican values and professional standards of practice that they all share at a broad level, and therefore ought to be granted some room to flourish under a system of carefully regulated autonomy and choice. Next I address three objections to a conception of school choice grounded in a notion of reasonable pluralism among educational doctrines: that it would undermine educator's efforts to secure status for themselves as professionals by admitting that "best practices" in education offer rough guidance at best, it would leave parents and students vulnerable to quackery, and that it abandons the common school tradition and its aspirations. I conclude with a brief discussion of why the conceptual basis on which a society designs a system of choice makes a difference, and why a basis in pedagogical pluralism has certain advantages more common bases of choice.

This argument differs from other pro-choice arguments in two ways. First, in linking education and morality, I am not speaking simply about religion, or about marginal cases where religious or ethnonationalist extremists create endless legal headaches for a dominant liberal and secular mainstream. The disagreements that concern me fall well within the mainstream of political and professional thought in the United States. The struggles between competing educational theories and methods over the last century and a half do not test the limits of liberalism or the Constitution in the same way that faith-based and ethnonationalist resistance to common schooling do. Cases involving religious minorities and state-operated public schools have been well-considered by others, including Rosemary Salomone (2000) and Stephen Macedo (2000). These treatments have addressed very real and intractable problems created by the presence of illiberal minorities under a liberal-democratic regime, and the philosophical and Constitutional questions they raise are of the utmost importance to liberal-republican theory and practice. But as applied to the questions of schooling and school choice, the focus on Constitutionally challenging cases can misleadingly suggest that there's a well-defined and articulated consensus in the US over what should be taught and how, and that those who challenge this consensus are somehow unreasonable or even threatening to the liberal-republican order that the rest of us seek to preserve. The first suggestion can make school choice seem superfluous or distracting. The second can make school choice sound dangerous, conjuring visions of publicly supported schools that preach hatred, oppression, or anti-Americanism. I want to focus instead on the fault lines within the loose liberal and professional consensus where these specters do not present themselves so acutely.

Second, my argument speaks primarily to educators themselves. It attempts to take seriously, and treat sympathetically, some of their deepest and most divisive professional convictions. Nearly the entire corpus of school choice literature focuses on why choice is good or bad for students, parents, and civil society. While I think my argument applies to all three, it is teachers who come closest to having articulated, informed, and deeply held beliefs about teaching and learning. And it is largely educators and their organizations that have fueled the "curriculum wars" of the 20th century. (Note 1) Education is their vocation, and educators' identities tend to be far more deeply conditioned by a given conception of that vocation than other constituents; they therefore have too much at stake in the outcome of their struggles. Howard Gardner (2000) and Deborah Meier (1995) have written suggestively in this area, Gardner acknowledging that competing conceptions of good schools might require accommodation and Meier suggesting that school choice might actually be liberating for teachers. Like theirs, mine is a pro-choice, pro-educator argument motivated by a desire to realize conditions under which educators can do their best work on behalf of children, families, communities, and the republic.

I think one of the reasons educators are so hostile to choice is that so many proponents of choice tend to treat public education as a monolithic establishment to be resisted, or reduce what educators tend to see as a moral project to a system of "service providers" catering to clients, or worse, customers. More strident commentators have construed choice as a way of breaking up unions or the "educational monopoly," or as an escape hatch for "underserved" families neglected by the uncaring monolith. This characterization cannot sit well with working educators, most of whom feel beleaguered and hamstrung by policymakers and each other in their efforts to do right by their students. A conception of school choice—and policy in general—that recognizes educators' role as moral agents in the formation of good persons restores some honor to the vocation and emphasizes some of the ways in which educators and their constituents are allied, rather than opposed. It is time to reframe the choice debate.

The Philosophical and Ideological Sources of Educational Doctrines

This section makes what I hope is an obvious point: educational doctrines are not mere preferences or prejudices, but are expressions of belief systems informed by deeper philosophical and ideological convictions. For purposes of argument it will be helpful to identify some broad categories of commonly recognized educational doctrines. Observers who have surveyed the 20th century have settled on roughly four. (Gutek, 1997; Kliebard, 1995; Pulliam, 1995; Partington, 1987). Their classifications vary somewhat, but they cite the same key figures, movements, and permutations, so that these can stand for a rough consensus on a classification scheme that captures reasonably well the landscape of competing educational visions in the 20th and 21st centuries. Roughly speaking, these taxonomists have identified two kinds of "traditionalism" and two kinds of "progressivism." Among traditionalists are those who uphold the humanistic and liberal arts model of education focused on high culture and generally (though not exclusively) grounded in Western intellectual traditions, and those believe that schools should inculcate skills, knowledge, and behaviors that will enable students to become productive worker-citizens. Though they differ from each other in important ways, both kinds of traditionalist favor what might be called academic learning—classroom-centered, text-based, and largely disciplinary. They also share a tendency to maintain more authoritative, formal relations between adults and students, and to insist on common standards of comportment defined by prevailing cultural norms of civility. Progressivism also comes in two broad varieties. The first is what proponents like to call "student-centered," which tries to organize learning around the talents and needs of each individual child. It tends to eschew the academicism of the traditionalists in favor of applied, hands-on, "real-world" experiences whose content is determined as much as possible by students' interests. The second type of progressivism seeks to use schools as instruments for reconstructing society by socializing students to what adherents regard as more just and humane social norms than are said to be held by traditionalists and the society they represent. These two forms of progressivism likewise have their differences, but share an approach to adult-child relations that stresses negotiation and child involvement in codetermining rules of civility and comportment while giving as much rein as possible to individual and subcultural self-expression. These thumbnail descriptions are broadly sketched and oversimplify a messier reality, but I trust they look familiar enough to informed readers.

In his unusually slim and readable textbook, *Philosophical and Ideological Perspectives on Education* (1997) Gerald Gutek does the most explicit job among the taxonomists of tracing each educational doctrine to its philosophical, ideological, and historical origins, and so it is Gutek's analysis I draw on here. What I've been loosely calling educational doctrine, Gutek calls educational *theory*, which is basically a set of normative beliefs about what should be taught and how that derives both from experience and its interaction with larger and more comprehensive bodies of thought. Each theory operates under certain philosophical assumptions about the nature of reality, human nature, human knowledge, and ethics, and carries with it certain assumptions about the nature of societies, their histories, and the experiences of people within them. They are, in other words, informed by philosophy and ideology. It is to Gutek's credit that he recognizes the distinction. (Note 2) Where philosophy consists of abstractions and statements about metaphysics, epistemology, axiology (ethics and aesthetics), and logic, ideologies are the concrete and specific belief systems of specific groups interpreting their past, assessing their present, and attempting to enhance their status in the future. Where philosophy attempts to be universal, transcendental, and contemplative, ideology is partisan, historical, and activist. Together, Gutek argues, the two interact in a variety of ways with each other and experience to generate educational theories.

So, for example, traditionalists are more likely to subscribe to elements of a realist epistemology that holds that humans *discover* how the world works through disciplined investigation and reason, whereas progressives hew closely to Dewey's claim that humans *construct* models of reality as they encounter and solve problems in their environment. Likewise, traditionalism tends to stress the fixed and universal dimensions of human nature, especially the human capacity for reason, where progressivism emphasizes human plasticity under varying environmental conditions. (Note 3) And where traditionalists tend to follow Aristotle in thinking of autonomy as something one earns through disciplined mastery of essential knowledge and skills whose standards of excellence are prior to the individual, progressives follow a post-Rousseauian model where a person becomes autonomous by exercising autonomy early and often, and by cultivating a distinctive persona.

Ideologically one finds similar patterns. Traditionalism, as the name advertises, has overtones of Burkean conservatism, where tradition is an important social glue and source of collective wisdom, and where effective social change must be carried out incrementally within the tradition to preserve continuity and social cohesion. Progressives, on the other hand, tend to subscribe to assorted 19th and 20th century ideologies that view inherited traditions as at best an encumbrance to social and intellectual improvement and at worst part of an oppressive apparatus of power wielded by elites seeking to dominate others.

Once again, this sketch is overdrawn, but recognizable. Needless to say, the alignment of particular philosophical and ideological frameworks with educational theories is not always as predictable as the rough sketches suggest. It is possible, for example, to be a postmodernist-traditionalist (e.g., Richard Rorty, 1992), a classicist-feminist-social reconstructionist (e.g., Martha Nussbaum, 1997), or a realist-progressive (e.g., Rousseau). Kliebard calls these "hybrids" (p. 179). Whatever the particular relationship, though, educational theory always and inevitably develops out of some broader philosophical and ideological frameworks. (Note 4)

The philosophical and ideological underpinnings of their educational theories and commitments are not always fully articulated or even recognized by adherents. In fact, most adherents of a particular set of doctrines most often regard them as "common sense," a straightforward description of the world (and ethics and knowledge) as it simply *is*, a state of affairs obvious to all but fools, knaves, and enemies of children. But the frameworks are their guiding thought and action nonetheless.

Whatever the particular configuration, educational theories reflect comprehensive worldviews and normative beliefs that profoundly condition persons' consciousness, particularly their conceptions of a good life and just society. As with all comprehensive doctrines, secular and sacred, different educational theories are mutually incommensurable. And yet, they represent disagreements among *reasonable* people who all subscribe in one way or another to some aspect of post-Enlightenment, post-Romantic liberal and republican thought. This acknowledgment is crucial. Most educational theories/doctrines in the US share certain aims. They all seek to cultivate tolerant, just, reasonable, critical-minded, and autonomous persons who are productive workers, competent and informed citizens, and adaptive agents able to negotiate a complex and changing social, political, and economic environment. It is just that they interpret these broadly shared ends through different philosophical and ideological filters that lead them to construe their educative mandates in sharply conflicting ways. All sides sincerely value republican citizenship and equality for all students. But does equality mean equal access to Euro-American high culture, as humanistic traditionalists maintain, the equal representation of the literatures of oppressed peoples, as social reconstructionists believe, or equal opportunity to read what one is most interested in or choose to study films instead, as a child-centered progressive might aver? Likewise, does educating for equal democratic citizenship require that we all become facile in the traditions and discourse of the civilization that gave rise to citizenship as we understand it, that each subcultural group have its own traditions and discursive modes be recognized and incorporated into the academic and social life of the school and public life more broadly, or that children begin making collective and individual decisions about fundamental matters of curriculum, comportment, and dress as early and often as possible? Each of these positions has plausible arguments on its side, and represents a reasonable interpretation of equality and citizenship. But they are irreconcilable with each other.

One of the beliefs that all sides share to some degree is a belief in the benefits of pluralism to a robust public culture. If so, then why not accept pedagogical pluralism for what it is, embrace it, and find a way to grant it fuller expression within reasonable bounds?

Common Schools and the Profession: Embodiment of "Best Practice" or Established Church?

In coming to grips with this diversity of educational doctrine, the first thing to recognize is that this pluralism is OK. In fact, it is natural. As Rawls put it, "A plurality of reasonable yet incompatible comprehensive doctrines is the normal result of the exercise of human reason within the framework of the free institutions of a constitutional democratic regime" (p. xviii). The proliferation of differing worldviews forms the warp and woof of life in a liberal republic, and suppression of dissent never works for long. So if pluralism of pedagogical doctrines reflects broader philosophical and ideological pluralism, the question becomes how best to manage it. Scholarly attempts to address pluralism in schools have focused on the cultural or religious identifications of students, parents and communities, and have largely been proposed within the paradigm of the common school. Crudely put, this scholarship can be cast as a debate between pluralists, who aver that schools should accommodate and even promote as broad an array of individual and group differences as fully as possible; and assimilationists, who argue that common schools should forge a common culture. In recent decades, the advantage has gone to the pluralists—"we are all multiculturalists now," as Nathan Glazer (1998) famously put it.

Paradoxically, the overwhelming cultural shift in favor of pluralism has been accompanied by an unabated attempt to impose a uniform educational theory (and practice) on the entire institution of schooling. Whether attempting to define national content standards or attempting to establish a uniform canon of "best practices," each group of advocates believes it is trying to do best by kids and society, and therefore must prevail. Yet experience and research have shown that students can thrive in a broad range of schools, from Core Knowledge schools and KIPP academies to women's leadership schools and expeditionary learning centers. All these types of schools produce graduates who are literate, productive citizens. As Larry Cuban (2000) observes, there are many different ways for a school to be "good." As we saw with respect to political and social values, beneath the doctrinal conflict there lies a set of characteristics that all good schools have in *common*:

> They have clear and shared purposes; they believe that all children can learn; each school staff has developed a working culture that embodies these common beliefs and enjoys collective action; and parents are deeply involved with the school. Thus very different concepts of schooling can be embraced without sacrificing the core purposes of public education. (p. 152)

Variations on Cuban's list of attributes are found widely in accounts of successful schools. At this level, educators enjoy notable consensus. For example, in one way or another, Theodore Sizer (1997), Mortimer Adler (1982), and Paul Gagnon (1993) have in different ways argued that "less is more"—i.e., that curricula should be selective, covered in-depth, and coherent. The problem is that the consensus erodes quickly as one begins to unpack their assertions: the criteria of selection, the meaning of depth, and the principles of coherence differ substantially among the progressive Sizer, the humanist-traditionalist Adler, and the disciplined-based traditionalist Gagnon. Where Sizer calls for thematic projects as the basis of curricular organization, Adler looks to the Great Conversation among classic Western authors, and Gagnon to the internal structures of the academic disciplines within the broad sweep of history. Though Sizer and Adler once collaborated for a time on the Paideia project in the 1980s, these three men could not together create and sustain a good, coherent school. Each by himself, in cooperation with like-minded colleagues, could.

The kinds of conflicts represented by these three figures play themselves out daily in schools and communities across the US. A great deal of the acrimony among educators, and the incoherence of American schooling generally, stem from adherents of incompatible doctrines being forced to compete with one another for dominance within the common school. In a typical school partisans coexist warily and resent each other's influence. The traditionalists lament the lack of rigor, the progressives complain about the amount of required content coverage, and the reconstructionists sneer at the traditionalists' easy absorption of women authors and black inventors. Everyone is dissatisfied.

How then to improve the likelihood that that educators in a school will develop a cohesive culture, collegial environment, and collective mission? A critical enabling condition for these qualities is professional autonomy within a system of choice. This is intuitive for many reasons, but the most relevant here is that one of the inhibitors of cohesive school cultures is disagreement over the best curriculum, instruction, and school culture. As noted, every zoned school is a mish-mash of progressives, social reconstructionists, and traditionalists, and each group further contains its own internal factions and fault lines. Genuine collegiality grounded in a set of shared standards of practice and shared normative understanding with regard to the work to be done is nearly impossible under these conditions. In a very real sense many teachers cannot even comprehend one another. One reason that comprehensive school reform is so difficult is that any attempt to impose coherence inevitably favors one educational doctrine over another, thereby galvanizing resistance among a plurality of faculty. Some resist out of sheer lassitude or intransigence, sure; but many do so because they sincerely hold contrary beliefs about their vocation as educators. To keep the peace, schools often resort to giving each teacher as much latitude as possible to do as he or she pleases—hence, the oft-cited isolation of teachers and the difficulty of forming genuinely collegial school cultures. Or alternatively, they engage in perennial rituals of "consensus-building," which paper over the differences by temporarily retreating to the level where the consensus Cuban speaks of is possible. If educators could instead form around a particular educational theory, a common definition of citizen, worker, and lifelong learner, one of the chief barriers to cohesion and collegiality would abate. And the way to do this is to allow teachers (and parents and students) to choose those that best match their own philosophical and pedagogical convictions. (Note 5)

The argument that school choice grounded in a recognition of reasonable pluralism could benefit educators and other school constituents is admittedly speculative. But there is a historical analogy that I think gives us some warrant for considering it. It is worth recalling how Western societies first came begrudgingly to accept pluralism as a fact of life and eventually to regard tolerance as a prime public value. Prior to the 17th century the notion that political authority and social order could exist without an established church was unthinkable. Clearly, secular authority required divine sanction, mediated through the offices of an established church representing a particular set of doctrines. It took the Protestant Reformation and over a century of bloody, destructive conflict to convince people that this belief and the policies that stemmed from it had become sources of political and social *in*stability. Disestablishment and policies of religious tolerance emerged as pragmatic accommodations to political reality, a way to discourage people from killing each other over doctrinal differences. It was only later that intellectuals began to theorize tolerance as a positive good, something valuable in its own right and worthy of the strongest protections. Over time, most religious sects thrived as voluntary associations, and eventually came to recognize their common aims and interests. Catholics and Protestants, Anglicans and Baptists—not to mention Buddhists and Muslims—now coexist amicably in a way unimaginable to their 17th century predecessors. And all but the most extreme sects contribute to, rather than threaten, the common civic culture.

I don't want to press the analogy too hard. For one thing, common schools could be said to have achieved a *modus vivendi* among different adherents, and dissidents enjoy tenure protections that amount to something approaching an official policy of toleration in schools already. More fundamentally, the state has nearly opposite responsibilities with respect to religion and education—it is proscribed from supporting the one and obliged to support the other. On the other hand, I am not the first to observe the parallels between theological and educational doctrine. (Note 6) Insofar as the analogy does hold, it suggests that we may have less to fear from educational disestablishment than from a continual struggle among adherents of different education theories to establish their "faith" through the vehicle of the common school. And I do think it holds, at least to a point. Sectarian warfare among Christians obscured a great deal that competing groups held in common *as Christians*. Once again, beneath sectarian differences in education lies a great deal that most educators (and parents and the public) hold in common. It is quite plausible that as disestablishment paved the way for interfaith cooperation among religious sects, giving more freedom to educators to practice their sectarian creeds within the limits of the broad liberal-republican and professional consensus we do enjoy could heal current rifts among professional educators and enable greater professional solidarity. It could also prove energizing, as it did for religious sects, which thrived once their proselytizing energies were set free. After all, who among contemporary secularists feel the proselytizing impulse more strongly than educators?

Three Objections: Professionalism, Quackery, and the Common School Tradition

The claim that choice might be good for educators might provoke at least three objections, reflecting legitimate concerns among educators and the broader public. First, this argument implicitly denies that teaching is a profession on par with medicine, which has proved far more successful in establishing a tight canon of professional knowledge and best practices. This will naturally raise concern among those attempting to raise the status of teaching by analogizing from medicine. Second, allowing a broader array of schools and practices could increase the risk of hucksterism or quackery—that is, it opens the door to crackpots and assorted extremists, putting families at risk and further undermining the credibility of teachers. And finally, school choice seems to repudiate the common school ideal on which the modern American public school system was built. The common school objection should matter to educators as educators for two reasons: first, because the civic dimension of schooling forms part of its moral dimension and *raison d'etre*, and second, because public support for schools is said to rest in part on the sense people have that schools serve the public good.

School choice grounded in the recognition that pedagogical questions are as normative as they are empirical may prove difficult for professionalization advocates to accept. Educators have sought recognition as true professionals for the better part of the last century, a claim that has rested on the validity of the assertion that educators and educational researchers possess a scientifically-based based professional expertise not available to laypersons. They have never succeeded in securing the legitimacy they desire, at least in part because the claim to be scientifically-based is spurious. And where the science is sound, the implications for teaching practice usually leave considerable latitude for practitioners. For example, research on learning demonstrates convincingly that the mind actively constructs knowledge through its interaction with its environment. These findings tell you something about how the mind works and point roughly to phenomena that teachers ought to keep in mind when planning and delivering lessons; however, it does not say anything about the relative advantages of "discovery learning" versus well-delivered lectures or other modalities. It only tells you that, whichever modality you choose, it needs to incorporate certain strategies and take certain characteristics of students into consideration (Hirsch, 1996; Bransford et al, 2000). (Note 7) Yet partisans of each modality claim the research for themselves, claiming implicitly or explicitly that the research discredits their rivals. This partisan appropriation of research hurts both the credibility of the research and the public reputation of educators by making the former look cooked and the latter half-baked. You just don't see this kind of persistent doctrinal warfare in mainstream medicine.

If these claims sound like the *hauteur* of a philosopher, consider the following statement by the National Research Council's Committee on Scientific Principles for Education Research:

> A more global implication of the role of values in education research concerns the extent to which research in education is truly akin to an engineering science. The question of why education has not produced the equivalent of a Salk vaccine is telling. After all, medical research is something of an engineering science in that it brings theoretical understanding in the life sciences to bear on solving the practical problems of prolonging life and reducing disease. Education research is similar, with the key difference that there is less consensus on the goal. Medical research often has clearer goals—for example, finding a cure for cancer. Because values are so deeply embedded in education in so many different ways, education researchers do not have a singular practical goal that drives their inquiry. (2002: 85)

None of this gainsays the legitimacy or value of educational research, or its relevance to practice. It simply urges modesty about claims that a given pedagogical practice is "research-based," as well as greater candor about the degree to which values-based convictions drive what one does in the classroom with (or without) the research.

Nor should any of this be taken to deny that teachers require special skills and knowledge to do their jobs well. Quite the contrary; teaching requires considerable knowledge and skill. It just does not narrow the field of "best" education practices to a point where educators can declare a single best, empirically verifiable educational theory. In short, the research tells us what adherents of a particular education theory must take into account if they want to succeed under the terms of their doctrines, but it still doesn't tell us which doctrine we should all subscribe to. Again, education simply isn't medicine.

But the medical model is not the only professional model available to educators. Journalists, clergymen, and tradesmen all enjoy legitimacy as experts, and considerable respect from non-practitioners. And all hew to certain standards of vocational practice that permit a wide range of legitimate variation. Respectable journalistic styles range widely, as do the ideologies that inform them—from *National Review* to *The Village Voice*, and from literary journalism to Gonzo. Yet all honor a similar code of ethics with respect to standards of veracity, confidentiality of sources, and so on; and all hew to certain canons of rhetoric and style. Likewise with clergymen, tradesmen, and other practitioners of honored vocations. Each has standards of practice, but standards defined flexibly enough to allow highly diverse approaches. These standards are nonetheless tight enough to enable us to discern masters from quacks. Even modestly discriminating readers recognize the difference between *The New Yorker* and *The Weekly World News* (a supermarket tabloid). It is possible, then, that educational professionalism conceived more modestly could actually *enhance* teachers' status by aiming for a more plausible standard of professional legitimacy; one that, like journalism or ministry, allows a range of approaches within a more parsimoniously defined set of standards where there *is* broad consensus among educators and between educators and informed constituencies.

Could a more flexible conception of professionalism nonetheless open the door to quackery? There are, after all, avid readers of *The Weekly World News*. Religious cults abound. Fly-by-night hustlers have cashed many a check just before the new roof collapsed. Worse, it is usually the least well informed who are most susceptible to quackery, and the children of the ill-informed who are most vulnerable. Society simply cannot tolerate the educational equivalent of *The Weekly World News* or the Branch Davidians. The individual and collective stakes are too high, especially for disadvantaged children.

These concerns may have force in an unregulated environment where persons share no core values to bind together and sustain a public culture, no rough consensus about desirable educational outcomes, no standards of justice or means to enforce them, or no way to judge good from bad teaching. But none of these background conditions obtains in the United States. Once again, rough consensus on these standards does exist, despite differences with respect to how they are to be defined, codified, and instantiated.

In the trade and ministerial vocations, professional standards are maintained internally, *and* laws exist to protect persons from libel, fraud, or outright abuse. The regulatory bar could be said to be pretty low, especially for journalists or ministries, but there is no reason it cannot be set higher in education, in recognition of the higher stakes attached to it. Choice proponents have long recognized that choice requires standards and accountability provisions to help families and communities make good decisions (Finn, 2002). Defining standards broadly enough to accommodate legitimate differences among sectarians without undermining their value *as standards* poses both political and technical challenges. Figuring out how to assess student achievement of standards poses similar problems. But they are surmountable. Surely it is possible to strike some satisfactory balance between the need for meaningful standards and the accommodation of reasonable differences.

The common school objection poses the strongest challenge to the move toward a system of choice grounded in a recognition of reasonable pluralism. The common school by definition militates against it. It is an honorable institution through which Americans have expressed their liberal-republican ideals for the better part of two centuries. In it, children of varied ethnic and social backgrounds are supposed to mix and mingle and emerge as a unified citizenry with a common civic identity and equal opportunity for future prosperity. Some latter-day champions of the common school have even suggested that the common school serves as a sort of training ground for citizenship by acting as a goad to local political engagement (Gutmann, 1999). A system of choice is said to undermine all these aims.

Choice proponents have responded to the equity and civic challenges by pointing out that common schooling as it operates in practice in fact produces profound inequities, and that private schools have done at least as good a job forging the kinds of citizens we say we want—law-abiding, tolerant, engaged—as public/common schools. This rebuttal has some force. The common school simply hasn't lived up to its promise as equalizer of opportunity or forger of competent citizens and, furthermore, a system of choice does not require us to abandon either ambition. Constitutional law, liberalism, and republicanism provide fairly robust guidelines for ensuring that schools serve certain public, collective purposes and forge shared civic values robust enough to maintain a liberal-republican polity. And while Macedo and Gutmann are certainly correct in arguing that many of the political and social virtues we take for granted—such as tolerance and willingness to work together—actually require active cultivation or "conscious social reproduction," there is no reason why baseline beliefs like these cannot form the criteria by which schools are to be approved, accredited, and evaluated. Even libertarian-leaning choice proponents have acknowledged that school choice does not preclude regulation to ensure that schools promote academic and civic standards (Moe, 2002).

But then the school choice critic can ask, with some plausibility, why we don't just draw on these resources to improve common schools? Choice advocates have an answer to this, too—that political control of schools will by its nature always frustrate the goals of reformers; a market-based system would deliver most of what we want from schools more efficiently (Chubb and Moe, 1990). I would like to take a similar, but less market-oriented approach and suggest that, especially in an age of assertive pluralism such as ours, the common school actually works against its own best intentions by embroiling schools in ceaseless conflicts over the same sorts of normative questions that inhibit collegiality among educators, thereby undermining cooperation between schools and their constituents. I aver that perhaps school choice conceived as accommodation of persons' reasonable differences with respect to pedagogical doctrines, could actually enhance civic comity among American subcultures in the same way it could enhance professional collegiality among American educators, while boosting support for public schooling.

The problem with the common school is that it doesn't seem to produce the civic outcomes it strives for. Doctrinal conflicts about everything from math curricula to dress codes continually factionalize and polarize the very persons who most need to work together to makes schools successful. Again, defenders of the common school argue that these conflicts are a good thing, because they represent direct democracy in action, from which students and adults alike learn how to be engaged citizens. What they tend not to note is that the process rarely generates satisfactory, consensual resolutions. Rather, disputes are decided by factional wrangling, power politics, and litigation. These means of adjudicating conflict have produced timorous, incoherent, mediocre schools and fractious, litigious school constituencies. These supposedly democratic practices have largely interfered with both good pedagogical practice and civic comity. In other words, the common school may have become a source of instability inadvertently subverting its own best intentions.

This claim makes sense when we pause to consider a basic precondition for citizen consent to state rule. As William Galston (2000) has recently rehearsed it, "Genuine civic unity rests on unforced consent. States that permit their citizens to live in ways that express their values are likely to enjoy widespread support, even gratitude. By contrast, state coercion is likely to produce dissent, resistance, and withdrawal" (p. 108). This truism holds at the local level, as well, where district policies and school practices inevitably alienate some group or another. Despite the best, most sincere intentions of school and district personnel, some constituents inevitably experience their actions as coercive. Conservative Christians are a good example. As Michael Apple *and Anita Oliver (1996) have documented, their militancy is often provoked by their marginalization within the* public school system. The same phenomenon has been observed among ethnic subcultures. Fears not just of unfair procedural treatment (e.g. disproportionately high assignments to low academic tracks), but also substantive fear of "deculturalization" and "linguistic genocide" have tended to galvanize ethnic self-assertion and resistance (Spring, 2000). (Note 8)

Similar alienation among professional educators (and parents) has given rise to a bevy of organized dissident groups, each of which feels itself oppressed by a dominant educational establishment. Members of the progressive Coalition of Essential Schools and traditionalist Core Knowledge Foundation, for example, each see themselves as virtuous minorities fighting the good fight against the educational establishment. And curiously, each sees each other and the point of view they represent as embodying all that's wrong with the establishment. Meanwhile, the field is rife with groups who militate for or against phonics instruction, multiculturalism, school uniforms, bilingual education, and so on, all galvanized by real or perceived slights by a real or perceived establishment. The proliferation of these groups ought to strike us as odd: When a profession is at war with itself over whether young children should receive direct and systematic instruction in textual decoding *or* be immersed in "literacy rich environments," something has gone terribly, dishearteningly wrong.

The liberal-republican state still has a prerogative ensure that its citizens are educated to achieve reasonable standards of intellectual competence, and to endorse understandings of justice, tolerance, and public spiritedness consistent with itself. It's just that the usual mechanisms for coping with the demands of pluralistic constituencies in common schools— factional wrangling, litigation, dilutive accommodation—have proved unsatisfactory to nearly everyone. A system of public school choice that recognized a diversity of goods with respect to what's worth knowing and how it is taught could defuse some of the acrimony and restore some of the coherence. The potential benefits are twofold. At the school level, it has the potential to enhance professionalism and collegiality among teachers by allowing them to form communities of practice around some core conception of the pedagogical good. This makes possible agreement on principles, practices, and strategies to guide the work of the school. It likewise provides a substantive basis for parent and student buy-in up-front. So right away, two key features of strong schools—quality teachers and engaged students and parents—can more easily gain a toehold.

At the community level, support for schools might actually be enhanced under a regime of public school choice, because fewer people would feel compromised, silenced, or alienated. Accommodation could certainly defuse a lot of conflict not just among educators, but between educators, families, and communities. Choice alone would not be enough. A spirit of tolerance would also need to be cultivated in localities so that certain kinds of schools were not prevented from opening because of local majority bias, which would only shift the current acrimony to slightly different terrain while continuing to hold children hostage to doctrinal zeal. If this spirit were achieved, however, choice could actually make it easier for local citizens to like each other and their schools, which would represent a significant step forward.

To make the conceptual shift to school choice, we don't have to give up our commitment to basic fairness, common civic culture, academic standards, or certain common features of schooling. Nor should we. We simply have to find better institutional mechanisms for realizing them. If, at a certain level of abstraction, we all believe in problem-solving, literacy, and life-long learning; cooperation, justice, freedom, republicanism, patriotism, and tolerance; active student learning, curricular coherence, and authentic assessment; but let our ourselves get bogged down in doctrinal disputes about what kind of school best honors these, then we are all perhaps better served by allowing a reasonable pluralism to prevail. Paradoxically, a policy of pedagogical disestablishment could diminish sectarian rivalry and pave the way for greater interfaith cooperation, to the benefit of the common good. (Note 9)

Conclusion

My argument is not intended to imply that educators should exercise sole, or even primary authority over the kinds of schools that will be offered. The emphasis on professional educators does recognize that educators (and parents/citizens-turned-educators) are more likely to have at least a semi-coherent and semi-articulate educational philosophy. They will largely determine the kind of schools available, and will continue to work to convert others to their pedagogical worldviews. The emphasis on pedagogy and philosophy also recognizes and honors the pluralism that exists among reasonable, well-intentioned educators. Nonetheless, parents, community members, and other educational constituents also have worldviews and interests that demand voice and accommodation. My argument is compatible with other values-based rationales for choice, such as those put forward by Salomone (2000) and Galston, which are grounded in parental values and children's differing needs. The eventual landscape of schools would over time be shaped by the ongoing negotiations among educators, families, and other constituents.

The pluralistic model of school choice grounded in educators' (as well as communities' and parents') philosophical commitments also has some compatibility with the market-based model. Student achievement and life outcomes would still be a chief criterion for judging school quality. Schools would, in effect, compete with one another for the loyalties of students and parents. But there are substantive differences, as well, with consequences for how we think about choice and frame education policy generally. The chief difference lies in the conceptualization of schooling itself. Where the market model tends to conceive schooling as a service commodity, with educators as "providers" and families (and businesses) as "clients" or "consumers," the model sketched here comes closer to John Davison Hunter (2000) and Robert J. Nash (1997) who conceive schools as moral communities. I think this model more accurately reflects both how constituents experience schooling and how we *should* conceive it. The problem with the market model is that it provides meager conceptual resources for schools (or their authorizers) to exercise legitimate normative authority or impose reasonable expectations on students, families, business, and other constituents—there are, after all, no consumer obligations, only rights. Social progressives as well as conservatives have good reason to preserve a conception of schooling that recognizes, affirms, and supports the formative mission of schooling.

At the same time, the pluralistic model in some sense gives more power to families in that it urges policymakers to consider a broader range of goods when ascertaining school quality than students' academic achievement. Charter proponents have been frustrated by how difficult it can be to close a low-performing charter school in the face of family and community protest. This attachment to low-performing charters—irrational from a market perspective—reflects the diversity of goods that schools provide for teachers, parents, students, and communities. By honoring these, a pluralistic model would make school accountability more complex—a potential downside from the point of view of school quality measured primarily in terms of test scores or college going rates, but a potential enhancement for those who believe there is more to schooling than academic achievement.

If my argument has merit, the next step will be to sketch policy implications. Suffice it to say for now that while I think it makes a substantive difference in how we frame and think through policy questions, it does not by itself solve any of the implementation challenges that other school choice models face. Academic and civic standards would still need to be established. (Do we permit schools that teach Afrocentrism or give "equal time" to creation science?) "Reasonable pluralism" would have to be defined. (Is it reasonable for a school's dress code to require girls to wear head scarves?) Assessment systems would need to be devised that preserved rigor without unfairly favoring some kinds of schools over others. Fair and adequate funding formulae would need to be developed. Public information and transportation systems would need to be established. Regulations and incentives would need to be crafted that ensured that *all* families and students—irrespective of special needs, home language, race, ethnicity, or income—had full access to high quality schools. Decisions would have to be made about where to draw the line on school features that seem designed to appeal primarily to a single ethnic or religious group. (Note 10) (Head scarves, creation science, and Afrocentric schools are unlikely to draw many students from outside certain very particular subcultures.) Labor and certification issues would need to be addressed, new organizational networks formed, new ways of delivering electives, sports, and extra-curriculars. And so on.

These and many other problems of principle and practice would remain to be solved. But if designed correctly, a system of choice that honored the convictions of educators (and other constituents) would take some of the most intractable issues off the table, especially those that touch us most closely—curriculum, pedagogy, and standards of personal comportment. The liberal arts school could coexist with the project-based school, the JROTC academy with the school for peace and social justice, the school that requires uniforms with the one that allows students to collectively renegotiate the dress code every six weeks. Once all sects feel secure in the practice of their faith, might better schools and stronger professional solidarity follow? We won't know unless we call a truce in the pedagogical holy wars.

References

Adler, Mortimer J. (1982). *The Paideia Proposal. An Educational Manifesto.* New York: Collier.

Bransford, John D., Ann L. Brown, and Rodney R. Cocking (2000). *How People Learn: Brain, Mind, Experience, and School.* Washington, DC: National Academy Press.

Bryk, Anthony S., Valerie E. Lee, and Peter B. Holland (1993). *Catholic Schools and the Common Good.* Cambridge, MA: Harvard University Press.

Chubb, John and Terry Moe (1990). *Politics, Markets, and America's Schools.* Brookings Institution Press.

Committee on Scientific Principles for Education Research. *Scientific Research in Education.* Washington, DC: National Academy Press, 2002.

Cotton, Kathleen (2001). *New Small Learning Communities: Findings from Recent Literature.* Portland, OR: Northwest Regional Educational Laboratory.

Cuban, Larry (2000). "Why is it so hard to get 'good' schools?" *Reconstructing the Common Good in Education: Coping with Intractable American Dilemmas.* Eds. Larry Cuban and Dorothy Shipps. Palo Alto, CA: Stanford University Press.

Cushman, Kathleen (1999). "How small schools increase student learning (and what large schools can do about it," *Principal*, 79 (2): 20-22.

Finn, Chester E. (2002) "Real accountability in K-12 education: the marriage of Ted and Alice," *School Accountability: An Assessment by the Koret Task Force on K-12 Education*. Eds. Williamson M. Evers and Herbert J. Walberg. Palo Alto, CA: Hoover Institution Press.

Gagnon, Paul (1993). "What should children learn?" *Atlantic Monthly*, Dec: 65-78.

Galston, William A. (2002). *Liberal Pluralism: The Implications of Value Pluralism for Political Theory and Practice*. New York: Cambridge University Press.

Gardner, Howard (2000). *The Disciplined Mind: Beyond Facts and Standardized Tests, the K-12 Education that Every Child Deserves*. New York: Penguin.

Glazer, Nathan (1998). *We Are All Multiculturalists Now*. Cambridge, MA: Harvard University Press.

Gutek, Gerald L. (1997). *Philosophical and Ideological Perspectives on Education*. 2nd ed. Boston: Allyn & Bacon.

Gutmann, Amy (1999). *Democratic Education*. 2nd ed. Princeton, NJ: Princeton University Press.

Hirsch, E.D., Jr. *The Schools We Need and Why We Don't Have Them*. New York: Doubleday.

Hunter, John Davison (2000). *The Death of Character: Moral Education in an Age Without Good or Evil*. New York: Basic Books.

Kliebard, Herbert. (1995). *The Struggle for the American Curriculum: 1893-1958*. New York: Routledge.

Macedo, Stephen (2000). *Diversity and Distrust: Civic Education in a Multicultural Democracy*. Cambridge, MA: Harvard University Press.

Meier, Deborah (1995) *The Power of their Ideas*. Boston: Beacon.

Meyer, John (2000). "Reflections on education as transcendence," *Reconstructing the Common Good in Education: Coping with Intractable American Dilemmas*. Eds. Larry Cuban and Dorothy Shipps. Palo Alto, CA: Stanford University Press.

Mitchell, Katharyne (2001). "Education for Democratic Citizenship: Transnationalism, Multiculturalism and the Limits of Liberalism," *Harvard Educational Review*, 71(1): 51-78.

Moe, Terry M. (2002). "The Structure of School Choice," *Choice with Equity*. Ed. Paul T. Hill. Palo Alto, CA: Hoover Institution Press.

Nash, Robert J. (1997). *Answering the Virtuecrats: A Moral Conversation on Character Education*. New York: Teachers College Press.

Nussbaum, Martha (1997). *Cultivating Humanity: A Classical Defense of Reform in Liberal Education*. Cambridge, MA: Harvard University Press.

Paris, David C. *Ideology & Educational Reform: Themes and Theories in Public Education*. Boulder, CO: Westview Press.

Partington, Geoffrey (1987). "The Disorientation of Western Education: When Progress Means Regress," *Encounter*. Jan: 5-15.

Powell, Arthur (1996). *Lessons from Privilege: The American Prep School Tradition*. Cambridge, MA: Harvard University Press.

Pulliam, John D. (1995) *History of Education in America*. 6th ed. Englewood Cliffs, NJ: Prentice-Hall.

Ravitch, Diane (2000). *Left Back: A Century of Failed School Reforms*. New York: Simon & Schuster.

Rawls, John (1993). *Political Liberalism*. New York: Columbia University Press.

Rorty, Richard (1992). "Two Cheers for the Cultural Left." *The Politics of Liberal Education*. Eds. Darryl J. Gless and Barbara Herrnstein Smith. Durham, NC: Duke University Press.

Salomone, Rosemary (2000). *Visions of Schooling: Conscience, Community, and Common Education*. New Haven, CT: Yale University Press.

Sizer, Theodore (1997). *Horace's Hope: What Works for the American High School*. New York: Mariner Books.

Spring, Joel H. (2000). *Deculturalization and the Struggle for Equality: A Brief History of the Education of Dominated Cultures in the United States*. McGraw-Hill.

Wasley, Pat, Michelle Fine, M. Gladden, N.E. Holland, S.P. King, E. Mosak, L.C. Powell (2000) *Small School, Great Strides: A Study of New Small Schools in Chicago*. New York: Bank Street College of Education.

About the Author

David J. Ferrero, Ed.D
Director of Evaluation and Policy Research, Education
Bill & Melinda Gates Foundation
PO Box 23350
Seattle, WA 98102
david@gatesfoundation.org
dferrero@attbi.com

David Ferrero is Director of Evaluation and Policy Research for the education division of the Bill & Melinda Gates Foundation, and a former high school teacher. His scholarship and other writing have lately focused on applications of contemporary Anglo-American political philosophy to questions of educational policy and practice.

Notes

1. See Kliebard (1995) and Ravitch (2000) for two of the more comprehensive accounts of this history.
2. Most commentators do not. They tend instead to regard the conflicts among educational theories as primarily "ideological." Among partisans the ideological label is slur, implying insidious motives on the other side in contrast with the benign and virtuous motives of one's own. It is always the other side that's ideologically motivated. In scholarly treatments, where ideology is recognized as informing all sides, there is a tendency to regard ideology as a kind of false consciousness, a filter that interferes with consensus-building (Paris, 1995). Even Kliebard (1995), whose historical examination of the conflicts among different educational theories is admirably even-handed, reduces the conflicts to "symbolic politics," something vaguely irrational and ultimately ineffectual.
3. This is illustrative, but overdrawn. Traditionalists recognize the influence of environment, and progressives the power of human reason. Nonetheless, differences in emphasis at the philosophical level lead to considerable divergence at the pedagogical and practical.
4. Despite appearances, origins, and the avowals of some adherents, no position on the square is innately more *politically* conservative or progressive than the others. Jesuit education, for example, tends to be pedagogically traditional; yet the social mission of the Jesuits is progressive. Likewise, multiculturalists tend to regard themselves as political progressives, and tend to ally with pedagogical progressives as well. But multiculturalism originates in the ethnonationalist desire to hold on to a reified cultural identity over and against a broader and more inclusive civic identity, which is characteristic of certain conservatives. Religious fundamentalists, who like the Jesuits tend to be pedagogically traditional, are simultaneously reactionary conservatives and social reconstructionists of a different stripe.
5. There is some suggestive literature on the possible benefits to learning of voluntary association within schools. See, for example, Bryk et al (1993), Powell (1996), and Hunter (2000).
6. John Meyer (2000) actually defines educational theories *as* religions, using a set of criteria similar to those Gutek uses to define philosophy and ideology. Cuban (2000) likewise treats educational theories as quasi-religions: "By World War I, these competing progressive and traditional ideologies constituted different faiths in the best way of raising children. . . . This century-long see-saw struggle of ideas is, then a much deeper religious conflict over what role schools should play in society writ large and, more specifically, how children should be schooled" (pp.156-7). He draws here on several historical studies by Tyack and Hansot tracing the origins of different educational theories to religious sources.
7. Nor does it settle the epistemological debate between realists and pragmatists, as the question remains open whether the mental constructs correspond to something about the way the world actually is or are simply useful fictions that help humans solve problems in specific contexts.
8. I don't mean to imply here that we should encourage a system of school choice based on religious or ethnic identification. At the same time, there is no denying that the pedagogical traditions and the philosophies and ideologies that inform them developed out of particular cultural milieux. In a culturally pluralistic society, especially one characterized by aggressive subcultural self-assertion, we could expect a degree of interaction between pedagogical creeds and cultural affiliations. Might pedagogical creeds be used as cover for cultural or religious ones? If so, might a system of choice predicated on differing conceptions of best curriculum and pedagogy provide ethnic, religious, or socioeconomic groups a pretext for self-segregation? It is certainly possible, but it isn't as inevitable as it might at first seem. For example, in a recent case in Vancouver, British Columbia, Hong Kong immigrants' traditionalist pedagogical values clashed with the local native majority whose values were more progressive. Despite the case study author's attempt to characterize the conflict as racially motivated, it is noteworthy that the immigrant' educational values—characterized as Confucian and "Chinese" in origin—converged with those of conservative Christians in the area (Mitchell, 2001). Meanwhile, contemporary progressives working in the urban core of cities such as Chicago, Providence, and New York are demonstrating that progressive pedagogy—once regarded as effective primarily with white, suburban, affluent students—can also work for disadvantaged urban minorities (Cotton, 2001; Wasley, 2000; Cushman, 1999). These examples suggest that the interactions among pedagogy, philosophy, ideology, and culture are complex and potentially serendipitous. They hardly point to school choice grounded in pedagogical pluralism as a panacea for segregation or balkanization; but they do suggest policy and recruitment strategies mitigating these risks through a diverse array of schools that appeal across ethnic and religious lines.
9. These are empirically testable assertions, but as choice proponents are fond of saying, we have to try

these policies in order to research their effects.

10. See footnote 8 above.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State Univeristy

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona

casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz Garcia (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil
(AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

PDF file of this article available:

This article has been retrieved 3610 times since September 1, 2003

Volume 11 Number 31 **September 1, 2003** **ISSN 1068-2341**

Performance Standards: Utility for Different Uses of Assessments

Robert L. Linn
University of Colorado at Boulder
&
National Center for Research on Evaluation, Standards, and Student Testing

Citation: Linn, R. L. (2003, September 1). Performance standards: Utility for different uses of assessments. *Education Policy Analysis Archives, 11*(31). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n31/.

Abstract

Performance standards are arguably one of the most controversial topics in educational measurement. There are uses of assessments such as licensure and certification where performance standards are essential. There are many other uses, however, where performance standards have been mandated or become the preferred method of reporting assessment results where the standards are not essential to the use. Distinctions between essential and nonessential uses of performance standards are discussed. It is argued that the insistence on reporting in terms of performance standards in situations where they are not essential has been more harmful than helpful. Variability in the definitions of proficient academic achievement by states for purposes of the No Child Left Behind Act of 2001 is discussed and it is argued that the variability is so great that characterizing achievement is meaningless. Illustrations of the great uncertainty in standards are provided.

Measurement specialists generally conceive of achievement as a continuum and they prefer to work with scale scores with many gradations rather than with a small number of categorical scores. It is recognized that there are a number of purposes for which the scores need to be lumped into a small number of categories that require the identification of one or more cut scores. Some leading measurement specialists, however, have suggested that it is best to avoid setting performance standards and associated cut scores if possible. For example, Shepard (1979) advised that it is best to "avoid setting standards whenever possible" (p. 67) and Green (981) concluded that "fixed cutting scores should be avoided whenever possible" (Green, 1981, p. 1005).

There obviously are some purposes where the identification of one or more cut scores cannot be avoided because they are essential to the use of a test. Tests used to make licensure and certification decisions, must have a cut score identified that will collapse the score scale into only two categories – pass and fail. In other situations, the scores are collapsed into 4 or 5 categories. The College Board, for example, converts a weighted combination of scores on the multiple-choice and the constructed-response sections of their Advanced Placement (AP) Examinations to a final grade that is reported on a five-point scale:

- 5 – extremely well qualified
- 4 – well qualified
- 3 – qualified
- 2 – possibly qualified
- 1 – no recommendation.

The pass-fail dichotomy is required for the decision to be made in the case of a licensure or certification test. The use of 5 categories on AP examinations, on the other hand, also supports dichotomous decisions about whether or not a student will receive college credit based on his or her AP grade, but allows colleges and universities to determine the grade required to be awarded credit.

Certification tests and AP examinations are just two of many situations where the primary use of test scores is to determine whether the test taker has met a performance standard. The performance of the person who barely met the standard is more similar to that of the person who barely failed to meet the standard than to that of someone who exceeded the standard by a comfortable margin. Indeed, due to measurement error, there is a substantial probability that the person who barely met the standard may be a false positive while the person who barely failed may be a false negative. The two could easily switch places if they took an alternate form of the test. Although such classification errors are of concern and should be minimized to the extent possible, they cannot be avoided altogether and there are legitimate practical reasons that require that a decision be made. Thus, Shepard's (1979) and Green's (1981) advice to avoid the use of a fixed standard or cut score cannot always be followed because standards are an essential element of the use of the test results. In the last 10 or 15 years, however, performance standards have been mandated and used with increasing frequency in situations where they are not essential to the use being made of test results.

Nonessential Uses of Performance Standards

Performance standards began to be introduced in uses of tests where they were not really essential as the result of the criterion-referenced testing movement that was spawned by Glaser's (1963) classic article. Ironically, Glaser's conceptualization of criterion-referenced measurement did not require the establishment of a fixed standard or cut score, but the use of cut scores to determine that a student either did or did not meet a performance standard became associated with criterion-referenced tests (Hambleton & Rogers, 1991; Linn, 1994). Glaser's later discussions of criterion-referenced testing (e.g. Glaser & Nitko, 1971) recognized the use of criterion in the sense of a fixed standard noting that "[a] second prevalent interpretation of the term criterion in achievement measurement concerns the imposition of an acceptable score magnitude as an index of achievement" (p. 653). Nonetheless, the setting of a fixed standard or cut score is not an essential to criterion-referenced measurement.

There are a number of reasons to question the wisdom of setting a performance standard for a test if the standard is not essential to the use of the test results. Green's (1981) desire to avoid setting fixed performance standards whenever possible was based on the recognition that a single item provides very little information by itself but "one item may well make the difference between passing and failing" (p. 1005). Others who have been critical of the use of performance standards have focused on limitations of the standards. Based on his review of standard setting methods, Glass (1978), for example, concluded that standards setters "cannot determine 'criterion levels' or standards other than arbitrarily. The consequences of arbitrary decisions are so varied that it is necessary either to reduce the arbitrariness, and hence the unpredictability or to abandon the search for criterion levels altogether in favor of ways of using test data that are less arbitrary, and hence safer" (p. 237). Although respondents to Glass's article noted that although standards are arbitrary in the sense that they are set judgmentally, they need not be capricious (Hambleton, 1978; Popham, 1978).

For a licensure test there is a clear context for standard setters who have a responsibility for thinking about the minimal skills required to protect the public from incompetent practitioners. The broader good of requiring some minimal level of performance on a test to be certified and therefore allowed to practice justifies the judgmentally established cut score. Moreover, the need to make certification decisions provides judges with a clear context for setting the standard. Standard setters may also have a clear sense of the proportion of candidates who have passed the examination in the past. Similarly, for an AP test there is a clear link between college grades assigned in courses for which credit may be awarded. Judgments regarding minimal acceptable performance when standards are set on tests that must be passed to graduate from high school or for promotion to the next grade have a more generalized context than the one that applies to setting standards for licensure examinations. The need to establish a cut score is essential to the use of a test as a high school graduation requirement and at least some of the consequences of the use are known to the standard setters who can weigh the potential benefits (e.g., restoring the meaning of a high school diploma or motivating greater effort by teachers and students, Mehrens & Cizek, 2001, p. 480) against the potential negative consequences (e.g., that the minimums will become the maximum or that more students will drop out of school, Mehrens & Cizek, 2001, p. 481).

In many instances the standards that are set are not used to make any pre-specified decisions about individual students. Instead the performance standards are used for reporting the performance of groups of students and for tracking progress of achievement for schools, states, or the nation. Examples include the setting of performance standards for the National Assessment of Educational Progress (NAEP) or a state assessment. The context that standard setters have had for setting a cut score to determine proficient, basic, or advanced performance for NAEP or a state assessment has, until recently, lacked any clear context other than a sense of an aspiration for high levels of student achievement.

Standards and Aspirations

Six broad educational goals, two of which concerned student achievement, were agreed to by Governors and the President at the Education Summit held in Charlottesville, Virginia in 1989. The National Educational Goals Panel was created and given the responsibility of monitoring progress toward the goals set at the Education Summit. Five years after the Charlottesville summit, the Goals 2000: Educate America Act of 1994 was signed into law (Public Law 103-227). Goals 2000 encouraged standards-based reporting of student achievement. As defined by a technical planning group for the Goals Panel, "performance standards specify 'how good is good enough'" (National Educational Goals Panel, 1993, p. 22). Unanswered is the question: good enough for what? It is clear that the performance standards are expected to be absolute rather than normative. (Although normative comparisons have been eschewed by many proponents of criterion-referenced tests and standards-based assessments, norms have considerable utility for providing comparisons of relative performance across content areas and even for a single measure are often more readily interpreted than are criterion-referenced reports or reports of results in terms of performance standards (see, for example, Hoover, 2003).) In keeping with the zeitgeist of the time, it is also clear that they were expected to be set at "world class" levels.

The high aspirations of 1989 Education Summit which were encouraged by Goals 2000 and the Goals Panel provided the primary context for the setting of standards on NAEP and on many state assessments. As might have been expected the performance standards set on NAEP on a number of state assessments were set at high levels. In support of the judgment that the performance standards were set at ambitious levels consider the fact that in 1990, the first year that NAEP results were reported in terms of achievement levels, the proficient level on the mathematics assessment corresponded to the 87th percentile for 4th grade students, the 85th percentile for 8th grade students, and 88th percentile for 12th grade students (see, for example, Braswell, Lutkus, Grigg, Santapau, Tay-Lim & Johnson, 2001). Linkages of NAEP to the Third International Mathematics and Science Study (TIMSS) grade 8 mathematics results which revealed that no country is anywhere close to having all of their students scoring at the proficient level or higher (Linn, 2000) also attest to conclusion that the NAEP achievement levels are set at very ambitious levels.

Performance standards set by a number of states for their state assessments during the past decade have also been set at quite high levels in many cases. The high levels that were set by many states are evident from the linkage of state assessments to NAEP (e.g., McLaughlin & Bandeira de Mello; 2002, see also Linn, in press, for a discussion of the McLaughlin & Bandiera de Mello results). It is not unusual for a state to have the proficient performance standard set at the 70th percentile or even higher. Since performance standards on NAEP and on a number of state assessments have, in most cases in the past, had no real consequences for students and there is no requirement for actually achieving the aspiration of having all students at the proficient level or above, one might conclude that there is no harm in having high standards. There may be reason for concern that reports that, say, less than half the students are proficient may paint an unduly negative picture for the public, but many would argue that it is good to have ambitious goals even if they are never achieved. "If you reach for the stars, you may not quite get them, but you won't come up with a handful of mud either" (Samuel Butler, as quoted by Applewhite, Evans, and Frothingham, 1992, p. 22).

It is one thing to set performance standards at the height of stars when there are no requirements of achieving them, but it is another matter altogether when there are serious sanctions for falling short such as those that have recently been put in place by the No Child Left Behind (NCLB) Act of 2001 (Public Law 107-110). NCLB sets the goal of having all children at the proficient level or higher in both reading/language arts and mathematics by 2013-2014. It also requires schools, districts, and states to meet adequate yearly progress (AYP) targets in intermediate years that would assure that they are track to having 100% of the students at the proficient level or above by 2013-2014. There are severe sanctions for schools that fall short of AYP targets for two or more years in a row.

The percentage of students who are at the proficient level or above for purposes of NCLB is to be determined by state assessments using performance standards established by each state. "Each State shall demonstrate that the State has adopted challenging academic content standards and challenging student academic achievement standards that will be used by the State" (P. L. 107-110, Section 1111(b)(1)(A)).

All states had to submit plans explaining how they were going to meet the accountability requirements of NCLB by January 31, 2003. States that were in process of introducing new assessments or that had not yet set performance standards will be setting standards in quite a different context than existed prior to the enactment of NCLB. In light of the new context provided by NCLB, it reasonable to expect that they will set the standards at less ambitious levels than they would have been set a couple of years earlier. The standards recently set by Texas for their new assessment, the Texas Assessment of Knowledge and Skills (TAKS), are consistent with the expectation that states may set their sights a little lower in the context of NCLB.

States that already had their assessments and performance standards in place prior to the enactment of NCLB faced a dilemma. They confronted the question of whether they should stay the course, recognizing that their performance standards were set at levels that are unrealistic for all children to achieve within the next 12 years. Or should they lower their performance standards and risk being accused of dumbing down their standards? Some states, e.g., Colorado and Louisiana redefined their performance levels for purposes on NCLB. Colorado, for example, has reported results on the Colorado Student Assessment Program (CSAP) in terms of four levels: unsatisfactory, partially proficient, proficient, and advanced. Colorado will continue to use all four levels for reporting to parents, schools, and districts. For purposes of NCLB, however, Colorado has collapsed the partially proficient and proficient levels into one level called proficient.

NCLB Starting Points for States

In order to track their AYP toward the goal of 100% proficient or above by 2013-2014, states have to define percentage proficient starting points. The starting point for each subject (reading/language arts and mathematics) is defined to be equal to the higher of the following two values: (1) the percentage of students in the lowest scoring subgroup who achieve at the proficient level or above and (2) "the school at the 20th percentile in the State, based on enrollment, among all schools ranked by the percentage of students at the proficient level" (P.L. 107-110, Sec. 111 (b)(2)(E)(ii)). In most cases the latter value will be the higher one and define the starting point.

Because states have their own assessments and set their own performance standards it should not be at all surprising that state NCLB starting points are quite variable. Some states are yet to define their performance standards and/or starting points and some states have expressed their starting points in terms of scale scores that are make comparisons difficult. Percentage proficient or above for reading/language arts starting points are available for 34 states at grades 4 and 8 (Olson, 2003). At grade 4, the starting percentages range from a low (i.e., most stringent) of 13.6% for California to a high (i.e., most lenient) of 77.5% for Colorado, with a median of 49.35%. At grade 8, the stating points range from a low of 13.6% to a high of 74.6% with a median of 46.2%. As at grade 4, California and Colorado define the extremes at grade 8. At grade 4, eight states have starting point percentages of 34% or less and eight states have starting points 64% or more. The corresponding percentages at grade 8 are 35% and 60%. State NAEP results indicate that states do vary in terms of student achievement, but not nearly enough to explain the huge variability in NCLB percentage proficient starting points. For the 43 states that participated in the 2002 NAEP 4th grade reading assessment, for example, the percentage of students who were at the proficient level or above ranged from a low of 14% in Mississippi to a high of 47% in Massachusetts (Grigg, Daane, Jin & Campbell, 2003).

The variability in the starting points is of similar magnitude for mathematics at grades 4 and 8 as that found for reading/language arts. The range for mathematics at grade 4 is from 8.3% in Missouri to 79.5% in Colorado and at grade 8 the range is from 7% in Arizona to 74.6% in North Carolina. On the 2000 NAEP mathematics assessment, North Carolina students did perform somewhat better than Arizona students. Thirty percent of the North Carolina students were at the proficient level or above on the grade 8 mathematics compared to 21% in Arizona (Braswell, Lutkus, Grigg, Santapau, Tay-Lim & Johnson, 2001). The grade 8 mathematics achievement of students in Arizona and North Carolina appears much more similar on NAEP, however, than is suggested by the starting points of 7% and 74.6%.

Controversy Regarding Performance Standards

Performance standards have been the subject of considerable controversy. The performance standards called achievement levels set on the NAEP) assessments, for example, have been subjected to harsh criticism. Reviews by panels of both the National Academy of Education (NAE) (Shepard, Glaser, Linn, & Bohrnstedt, 1993) and the National Research Council (NRC) (Pellegrino, Jones, & Mitchell, 1998) concluded that the procedure used to set the achievement levels was "fundamentally flawed" (Shepard, et al., 1993, p. xxii; Pellegrino, et al., 1998, p. 182). The conclusions of the NAE and NRC panels were controversial and several highly-regarded measurement experts have defended the procedure used to set the NAEP achievement levels as well as the resulting levels (e.g., Cizek, 1993; Kane, 1993, 1995; Mehrens, 1995).

There is an abundance of methods for setting standards, but there is no agreed upon best method. This point was made repeatedly at the Joint Conference on Standard Setting for Large-Scale Assessments sponsored by the National Assessment Governing Board and the National Center for Education Statistics held October 5-7, 1994 in Washington, DC. The Joint Conference included 18 presentations by scholars representing multiple perspectives. The papers dealt with a variety of issues ranging from technical to policy and legal issues. Crocker and Zieky (1995) prepared an executive summary of the conference which included the following summary conclusion.

"Even though controversies and disagreements abounded at the conference, there were some areas of general agreement. Authors agreed that setting standards was a difficult, judgmental task and that procedures used were likely to disagree with one another. There was clear agreement that the judges employed in the process must be well trained and knowledgeable, represent diverse perspectives, and that their work should be well documented" (p. ES-13).

Variability in Standards

As was indicated by Crocker and Zieky, there is a broad consensus in the field that different methods of setting standards will yield different standards. This consensus is consistent with Jaeger's (1989) summary of the literature on the comparability of standard setting methods. "Different standard-setting procedures generally produce markedly different standards when applied to the same test, either by the same judges or by randomly parallel samples of judges" (Jaeger, 1989, p. 497). It is also agreed that different groups of judges will set different standards when using the same method, especially when the groups represent different constituencies (e.g., teachers, administrators, parents, the business community, or the general public). Moreover, there is general agreement that "... there is NO 'true' standard that the application of the right method, in the right way, with enough people, will find" (Zieky, 1994, p. 29).

Given that there is no "true standard or "best" method for setting a standard, it is reasonable to ask what should be made of the variability in results as the function of choice of methods or choice of judges. If one wants to make generalizations across methods or groups of judges then it would seem reasonable to treat the variability in results as error variance. In doing so, we would at least acknowledge that there is a high degree of uncertainty associated with any performance standard.

Variability Due to Judges

Attempts are often made to estimate the error variability due to judges as part of the standard setting process. A difficulty that is encountered, however, is that standard setting methods usually involve group discussion following an initial set of judgments which may be made independently. Group discussion obviously makes judgments in subsequent rounds dependent which makes it impossible to estimate the error variability due to judges in the final round of judgments. Furthermore, there are good reasons to believe that the person who leads the standard setting exercise may have an important influence on the outcome. Thus, what would be desired is something akin to duplicate-construction experiment that Cronbach (1971) proposed as a way to evaluate the content validity of a test. The duplicate-construction experiment would require that two teams of "equally competent writers and reviewers" (p. 456) independently construct alternate tests. The parallel in standard setting would involve the use of independent panels of comparably qualified judges set the standards under the direction of equally competent leaders using the same method and instructions. The variance in the standards for the two panels would provide an estimate of the amount of error due to the panel of judges and standard setting leader.

Since the bigger sources of variability in standards is apt to come from the way in which judges are identified and the method that is used to set standard, even the parallel of Cronbach's duplicate-construction experiment would greatly underestimate the real degree of uncertainty in the standards. Some idea of the degree of variability due to the identification of judges is provided by results of a study conducted by Jaeger, Cole, Irwin, and Pratto (1980). Jaeger and his colleagues had three panels, consisting of samples of teachers, school administrators, and counselors, respectively, independently set passing standards on a North Carolina test. The differences in the standard set by the different panels can be gauged by the magnitude of the differences in the proportion of students who would have failed the test according to the different groups of judges. On the reading test the proportion who would have failed ranged from a low of 9% to a high of 30%. The variability in failure rates was even greater for the mathematics test, ranging from a low of 14.4% to a high of 71.1%.

Variability Due to Method

Variability due to choice of method can be evaluated based on results of several different studies that were reviewed by Jaeger (1989). One of those studies where multiple methods were used, for example, was conducted by Poggio, Glassnapp and Eros (1981). They had independent samples of teachers set standards using one of four different standard setting methods: the Angoff (1971) method, the Ebel (1972) method, the Nedelsky (1954) method, or the contrasting groups (see, for example, Jaeger, 1989) method. Teachers set standard for tests at grades 2, 4, 6, 8, and 11. There was substantial variability in the standards set by the four different methods at every grade. At grade 8, for example, the four different methods would set the minimum passing score on a 60 item reading test at 28, 39, 43, and 48 items correct. If the most lenient standard had been used, just over 2% of the students would have failed whereas approximately 29% would have failed if the most stringent standard had been used.

In his summary of 32 contrasts of standards set by different methods Jaeger (1989) found that the ratios of the percentages of examinees who would fail range from a low of 1.00 to a high of 29.75 with a median of 2.74. That is, the typical consequence of using a different method to set standards would be to alter the failure rate by a factor of almost 3. As Jaeger (1989) concluded the "choice of a standard setting method is critical" (p. 500). He went on to endorse earlier suggestions by Hambleton (1980), Koffler (1980) and Shepard (1980; 1984) that "it might be prudent to use several methods in any given study and then consider all the results, together with extra-statistical factors, when determining a final cutoff score" (Jaeger, 1989, p. 500).

Because of the practical cost considerations, the use of multiple methods as input to a final standard setting decision is rare in operational practice, but that was the approach recently taken in Kentucky for the assessments introduced in the state in 2000. As was recently reported by Green, Trimble and Lewis (2003), the Kentucky Department of Education used multiple methods as input to their final standard setting when the state introduced a new testing system, the Kentucky Core Content Test (KCCT) in 2000. First, three different methods, the bookmark procedure (Lewis, Green, Mitzel, Baum & Patz, 1998; Mitzel, Lewis, Patz, & Green 2001), the Jaeger-Mills procedure (Jaeger & Mills, 2001), and the contrasting group (see, for example, Jaeger, 1989) were used to set cut scores to distinguish four levels of performance (novice, apprentice, proficient and distinguished) on each of 18 different tests used in the KCCT system for various grade levels and content areas. The results of the bookmark, the Jaeger-Mills, and the contrasting groups standards setting efforts were input to a synthesis process where the results were considered by teacher committees that recommended cut scores to the Kentucky State Board of Education (Green, Trimble, & Lewis, 2003; see also CTB/McGraw-Hill, 2001; and Kentucky Department of Education, 2001, for more detailed descriptions).

Table 1 displays the percentage of students at the proficient level or above on each of six KCCT subject area tests administered at elementary school grades according to the three standard setting methods. Also shown are summary statistics across the three methods, mean, standard error, minimum, maximum, and range as well as the percentage proficient or above according to the standard set by the synthesis panel. The standard error is simply the standard deviation of the percentages for the three different standard setting methods since the standard deviation may be interpreted as a standard error if the goal is go generalize across standard setting methods. As can be seen, the standard errors are quite large, indicating that there is considerable uncertainty about the percentage of students proficient or above due to standard setting method.

Table 1

Percentage of Students at the Proficient Level or Above on KCCT For Elementary School Grade Tests According to Standard Setting Method

Method or Statistic	Subject						Mean
	Reading	Mathematics	Science	Social Studies	Arts & Humanities	Practical Living/ Vocational Studies	
Bookmark	56.5%	35.2%	35.4%	48.4%	15.3%	44.8%	39.3%
Jaeger-Mills	15.3	20.7	4.7	4.5	11.0	16.4	12.1
Contrasting Groups	29.4	19.5	24.5	27.2	24.8	24.7	25.0
Methods Mean	33.7	25.1	21.5	26.7	17.0	28.6	25.5
Standard Error	20.94	8.74	15.56	21.95	7.06	14.60	14.81
Maximum	56.5	35.2	35.4	48.4	24.8	44.8	40.85
Minimum	15.3	19.5	4.7	4.5	11.0	16.4	11.9
Range	41.2	15.7	30.7	43.9	13.8	28.4	29.0
Synthesis Standard	57.2	31.2	35.9	39.8	13.3	45.4	37.1

Tables 2 and 3 display results parallel to those in Table 1 for the KCCT tests administered at the middle school and high school grades, respectively. The results for the 12 subject area by grade combinations shown in Tables 2 and 3 are similar to those shown in Table 1 for the tests administered at the elementary school grades. The mean standard error across the 18 subject area by grade combinations in Tables 1 through 3 is 10.82 and they range from a low of 4.26 for the middle school mathematics test to a high of 26.35 for the middle school reading test. Even in the best case there is a good deal of uncertainty about the percentage of students who are at the proficient level or above as the consequence of the method used to set the performance standards. A standard error of even 4 points is large relative to the annual change in percentage proficient or above that is likely to be required to meet the AYP target for NCLB. A standard error of 26 points is gigantic in that same context.

Table 2
Percentage of Students at the Proficient Level or Above on KCCT For Middle School Grade Tests According to Standard Setting Method

Method or Statistic	Subject						Mean
	Reading	Mathematics	Science	Social Studies	Arts & Humanities	Practical Living/ Vocational Studies	
Bookmark	61.0%	19.8%	50.6%	28.6%	41.6%	40.6%	40.4%
Jaeger-Mills	10.5	17.7	10.4	12.8	14.6	16.2	13.7
Contrasting Groups	22.7	25.9	23.7	22.8	23.9	30.9	25.0
Methods Mean	31.4	21.1	28.2	21.4	26.7	29.2	26.4
Standard Error	26.35	4.26	20.48	7.99	13.72	12.29	14.18
Maximum	61.0	25.9	50.6	28.6	41.6	40.6	41.4
Minimum	10.5	17.7	10.4	12.8	14.6	16.2	13.7
Range	50.5	8.2	40.2	15.8	27.0	24.4	27.7
Synthesis Standard	51.0	25.2	27.3	28.3	35.9	35.3	33.8

Table 3
Percentage of Students at the Proficient Level or Above on KCCT For High School Grade Tests According to Standard Setting Method

Method or Statistic	Subject						Mean
	Reading	Mathematics	Science	Social Studies	Arts & Humanities	Practical Living/ Vocational Studies	
Bookmark	27.1%	10.9%	19.0%	22.5%	21.0%	47.8%	30.7%
Jaeger-Mills	12.0	10.9	2.6	14.4	11.0	17.5	11.4
Contrasting Groups	26.3	25.9	30.9	24.0	24.5	33.5	27.5
Methods Mean	21.8	15.9	17.5	20.3	18.8	32.9	21.2
Standard Error	8.50	8.66	14.21	5.16	7.00	15.16	9.78
Maximum	27.1	25.9	30.9	24.0	24.5	47.8	30.3
Minimum	12.0	10.9	2.6	14.4	11.0	17.5	11.4
Range	15.1	15.0	28.3	9.6	13.5	30.3	18.6
Synthesis Standard	27.5	26.3	27.3	24.0	19.5	48.4	28.8

Of course Kentucky did not use any of the three methods to set the performance standards for operational use. Rather the results of the three methods were used as input to the synthesis panels that provided the final recommendations to the State Board of Education. Hence, it might be argued that the variability due to method is not relevant to judging the uncertainty of the final performance standards. Thus, a reasonable question is what is the degree of uncertainty for operational standards in Kentucky or any other state? Results recently reported by Hoover (2003) are relevant to this question. Hoover compared of the percentage of students labeled proficient or advanced according to three national test batteries and NAEP. Hoover's results show that performance standards that are finally adopted after much care and expense by national test publishers for their tests or by the National Assessment Governing Board for NAEP also have a great deal of uncertainty.

According to the three national tests the percentage of grade 4 students who are proficient or above in reading is 24% according to one test publisher, 40% according to another publisher, and 55% according to the third publisher. According to NAEP the figure is 31%. For grade 4 mathematics, the corresponding four numbers are 15%, 34%, 44%, and 26% (Hoover, 2003, p. 11). Hoover's comparisons also show that, whereas there is apparently a substantial decline in the percentage of students who are proficient or above from grade 8 to 9 (33% vs. 11%) according to one publisher, there is no such decline according to the other two publishers. While one publisher's performance standards show a fairly steady decline in percentage of students who are proficient or advanced in mathematics from grades 1 through 12 (from 41% to 5%) another publisher's performance standards show that slightly more than twice as many students (27% vs. 12%) are proficient or advanced at grade 12 than at grade 1.

Conclusion

The variability in the percentage of students who are labeled proficient or above due to the context in which the standards are set, the choice of judges, and the choice of method to set the standards is, in each instance, so large that the term proficient becomes meaningless. Insistence on standards-based reporting of achievement test results where such reporting serves no essential purpose is more harmful than helpful. This is particularly true in the context of NCLB where schools, districts, and states are subject to substantial sanctions based on the progress that is made against arbitrary performance standards that lack any semblance of comparability from state to state.

Several years ago I (Linn, 1995) described four uses of performance standards: exhortation, exemplification of goals, accountability, and the certification of student achievement. Performance standards and associated cut scores are essential only for the fourth use. Although reporting results in terms of performance standards is often done to exhort teachers and students to exert greater effort standard-based reporting is not essential to that use. Nor are performance standards essential for the purpose of exemplifying goals. NCLB and a number of state accountability systems depend on the performance standards, but that would not have to be the case.

One of the purposes of introducing performance standards is to provide a means of reporting results in a way that is more meaningful than a scale score. Certainly, reporting that a student performed at the proficient level appears more understandable than saying that the student has a scale score of 215. Parents familiar with student performance in school in terms of grades of, say, A, B, C, D and F might naturally assume that proficient is like a grade of B or C. Given the huge inconsistencies in definitions of proficient achievement and in the associated stringency of cut scores, however, it seems clear that attaching such an interpretation to performance levels cannot mean the same thing across states where standards vary so radically in their stringency. It would be better to find another way of dealing with these non-essential uses of performance standards and cut scores.

There obviously is a legitimate interest in being able to measure progress in student achievement. There are many ways of measuring progress and setting AYP targets that do not depend upon the reporting of results in terms of performance standards. Effect-size statistics that would measure the year-to-year difference in mean achievement scores in terms of the standard deviation of scores in the base year is one obvious way that progress could be measured. Holland's (2002) proposal to measure progress in student achievement by comparing cumulative distribution functions is another approach. Comparisons of cumulative distribution functions provide a means of monitoring changes in student performance throughout the range of performance. Changes in the percentage of student exceeding any selected score level can be readily determined rather than just focusing on one arbitrary cut score that corresponds to the proficient performance standard.

Comparisons might also be made to norms for a base or reference year. If improvement in performance in State A was large enough that three quarters of the students in 2013-2014 performed above the median level in 2002-2003 that would represent a large improvement in student achievement. It would also be readily understood that that students in state A generally had better achievement than students in state B where 55% of the students in 2013-2014 scored above the 2002-2003 median. Furthermore, the norm-based results for states A and B would be much more interpretable than a statement that three-fourths of the students in State C were proficient or above in 2013-2014 compared to only 55% in State D, because the meaning of proficient might be radically different for States C and D. Indeed, if the stringency of the proficient performance standard were as variable from state to state as it is now, it might well be that the achievement of students in state D was actually better than that in state C.

Finally, if it is decided that the best way to track progress is in terms of the percentage of students scoring above a fixed cut score, sometimes referred to as PAC for percent above cut, then it would be better to pick the cut score based on norms in a base year than to use an arbitrary definition of "proficient" performance that bears little similarity to the definition of proficient performance in another state. The median or some other percentile rank in a base year might be used as the cut score. This would provide a clear and consistent meaning that does not seem possible to achieve for the proficient performance standard. Using PAC statistics to track progress would provide a reasonable alternative to tracking progress in achievement for different states in terms of percent proficient or above. PAC statistics based on a cut score at, say, the median achievement level in a base year would also be much more interpretable than percentages proficient or above when comparisons are made across states.

Reports of individual student assessment results in terms of norms have more consistent meaning across different assessments than reports in terms of proficiency levels based on uncertain standards.

Furthermore criterion-referenced reports of results can be provided by illustrating the types of items that the student can answer correctly and the types that they cannot. A fixed cut score is not essential to criterion-referenced interpretations of achievement.

Postscript

For the reasons discussed above, I believe it would be desirable to shift away from standards-based reporting for uses where performance standards are not an essential part of the test use. I recognize, however, that existing state and federal laws require the setting of performance standards and the reporting of results in terms of those standards. Thus, at least until the laws are changed, there is no choice but to work to make performance standards as reasonable as possible. Assuring that judges on standard setting panels understand the context in which the standards will be used is a minimal requirement for obtaining reasonable performance standards. Normative information needs to be made part of the process for judges to anchor their absolute judgments with an understanding of current levels of performance of students and likely consequences. As Zieky (2001) has noted, considering both absolute and normative information "in setting a cutscore can help avoid the establishment of unreasonably high or low values" (p. 38). In addition to knowing the percentile rank corresponding to particular cut scores, it would also be desirable to have some means of providing judges with information comparative information about the relative stringency of their standards in comparison to standards set in other states before judgments are finalized. Normative information would be one way of making comparisons to standards in other states.

It is critical that the context in which the standards will be used be made as clear as possible to panels of judges who set the standards. The uses of the standards for purposes of NCLB with its expectation that all students will be at the proficient level or higher by 2014 and sanctions for schools that do not meet AYP targets is an important part of the current context that standard setters need to consider.

Finally, while there is no agreed upon best method for setting standards, the literature does provide useful indications of the differences among different methods in their relative stringency and ease of use. Jaeger's (1989) advice that multiple standard setting methods be used and the results of the different methods be "considered together with extrastatistical factors when determining the cutoff score" (p. 500) seems as sound now as it in 1989. The experience in Kentucky with the use of multiple methods as input to a synthesis panel provides an excellent example where that advice was taken seriously in practice.

A number of authors have suggested helpful criteria to consider in selecting a method. Hambleton (2001), for example, identifies 20 criteria that he presents in the form of questions to be considered in evaluating a standard-setting process. Raymond and Reid (2001) provide useful advice on the selection and training of judges. Kane (2001) provides a good discussion of considerations in the validation of performance standards and cut scores, and Bond (1995) has identified five principles intended to help ensure fairness. Although such considerations cannot eliminate the arbitrariness, they can help make the standards and cut scores more reasonable and more defensible.

Acknowledgment

Based on a paper presented at the College Board, New York City, July 16, 2003. Work on the paper was partially supported under the Educational Research and Development Center Program PR/ Award #R305B60002, as administered by the Institute of Education Sciences, U.S. Department of Education. The findings and opinions expressed in this paper do not reflect the position or policies of the National Center for Education Research, the Institute of Education Sciences, or the U.S. Department of Education.

References

Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.

Applewhite, A., Evens, W. R. III, & Frothingham, A. (1992). *And I quote*. New York: St. Martin's Press.

Bond, L. (1995). Ensuring fairness in the setting of performance standards. In *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments*. (pp. 311-324). Washington, DC: National Assessment Governing Board and National Center for Education Statistics.

Braswell, J. S., Lutkus, A. D., Grigg, W. S., Santapau, S. L., Tay-lim, B.S.-H., & Johnson, M. S. (2001). *The nation's report card: Mathematics 2000*. Washington, DC: National Center for Education Statistics.

Cizek, G. J. (1993). *Reactions to National Academy of Education report, Setting performance standards for student achievement*. Washington, DC: National Assessment Governing Board.

Crocker, L. & Zieky, M. (1995). Executive Summary: Joint conference on setting standards for large-scale assessments. In *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments*. (pp. ES-1 – ES-17). Washington, DC.

National Assessment Governing Board and National Center for Education Statistics.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507).

CTB/McGraw-Hill. (2001). *Kentucky standard setting technical report.* Monterey, CA: Author.

Ebel, R. L. (1972). *Essentials of educational measurement.* Englewood Cliffs, NJ: Prentice-Hall.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist, 18*, 519-521.

Glaser R. & Nitko, A. J. (1971). Measurement in learning and instruction. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 625-670). Washington, DC: American Council on Education.

Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement, 15*, 237-261.

Goals 2000: Educate America Act of 1994, Public Law 103-227, Sec. 1 et seq. 108 Stat. 125 (1994).

Green, B. F. (1981). A primer of testing. *American Psychologist, 36*, 1001-1011.

Green, D. R., Trimble, C. Scott, & Lewis, D. M. (2003). Interpreting the results of three different standard-setting procedures. *Educational Measurement: Issues and Practice, 22*, No. 1, 22-32.

Grigg, W. S., Daane, M. C., Jin, Y. & Campbell, J. R. (2003). *The nation's report card: Reading 2002.* Washington, DC: National Center for Education Statistics.

Hambleton, R. K. (1978). On the use of cut-off scores with criterion referenced tests in instructional settings. *Journal of Educational Measurement, 15*, 277-290.

Hambleton, R. K. (1980). Test score validity and standard-setting methods. In R. A. Berk (Ed.), *Criterion-referenced testing: Motives, models, measures and consequences* (pp. 80-123). Baltimore, MD: Johns Hopkins University Press.

Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 89-116). Mahwah, NJ: Lawrence Erlbaum.

Hambleton, R. K. & Rogers, J. H. (1991). Advances in criterion-referenced measurement. In R. K Hambleton & J. N Zaal (Eds.), *Advances in educational and psychological testing* (pp. 3-44). Boston: Kluwer Academic.

Hoover, H. D. (2003). Some common misconceptions about testing. *Educational Measurement: Issues and Practice, 22*, No. 1, 5-14.

Holland, P. (2002). Measuring progress in student achievement: Changes in scores and score gaps over time. *Report of the Ad Hoc Committee on Confirming Test Results.* Washington, DC: National Assessment Governing Board.

Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (ed.), *Educational Measurement* (3rd ed., pp. 485-514). New York: Macmillan.

Jaeger, R. M., Cole, J., Irwin, D. M., and Pratto, D. J. (1980). *An interactive structure judgment process for setting passing scores on competency tests applied to the North Carolina high school competency tests in reading and mathematics.* Greensboro, NC: Center for Education Research and Evaluation, University of North Carolina at Greensboro.

Jaeger, R. M. and Mills, C. (2001). An integrated judgment procedure for setting standards on complex, large-scale assessments. In G. J. Cizek (Ed.), *Setting performance standards, Concepts, methods, and perspectives* (pp. 313-318). Mahwah, NJ: Lawrence Erlbaum.

Kane, M. (1993). *Comments on the NAE evaluation of NAGB achievement levels.* Washington, DC: National Assessment Governing Board.

Kane, M. (1995). Examinee-centered and task-centered standard setting. In *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments.* (pp. 119-141). Washington, DC: National Assessment Governing Board and National Center for Education Statistics.

Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 53-88). Mahwah, NJ: Lawrence Erlbaum.

Kentucky Department of Education. (2001). *Standard setting: Synthesis of three procedures – procedures and findings.* Frankfort, KY: Kentucky Department of Education. Available at www.kde.state.ky.us.

Koffler, S.L. (1980). A comparison of approaches for setting proficiency standards. *Journal of Educational Measurement, 17*, 167-178.

Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J. (1998). *The Bookmark standard setting procedure: Methodology and recent implementations.* Paper presented at the 1998 annual meeting of the National Council on Measurement, San Diego, CA.

Linn, R. L. (1994). Criterion-referenced measurement: A valuable perspective clouded by surplus meaning. *Educational Measurement: Issues and Practice, 13* (4), 12-14.

Linn, R. L. (1995). The likely impact of performance standards as a function of uses: From rhetoric to sanction. In *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments.* (pp. 267-276). Washington, DC: National Assessment Governing Board and National Center for Education Statistics.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4-16.

Linn, R. L. (in press). Accountability: Responsibility and reasonable expectations. *Educational Researcher.*

McLaughlin, D. & Bandeira de Mello, V. (2002). *Comparison of state elementary school mathematics achievement standards using NAEP 2000.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April.

Mehrens, W. A. (1995). Methodological issues in standard setting for educational exam. In *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments.* (pp. 221-263). Washington, DC: National Assessment Governing Board and National Center for Education Statistics.

Mehrens, W. A. & Cizek, G. J. (2001). Standard setting and the public good: Benefits accrued and anticipated. In G. J Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 477-485). Mahwah, NJ: Lawrence Erlbaum.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. J Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 249-281). Mahwah, NJ: Lawrence Erlbaum.

National Educational Goals Panel. (1993). *Report of the Goals 3 and 4 Technical Planning Group on the Review of Education Standards.* Washington, DC: Author.

No Child Left Behind Act of 2001. Public Law 107-110.

Olson, L. (2003). "Approved" is relative term for Ed. Dep.: 11 states fully meet ESEA requirements. *Education Week,* Vol. XXII, No. 43. August 6, pp. 1, 34-36.

Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (Eds.). (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress.* Washington, DC: National Academy Press.

Poggio, J. P., Glasnapp, D. R., & Eros, D. S. (1981). *An empirical investigation of the Angoff, Ebel, and Nedelsky standard setting methods.* Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April.

Popham, W. J. (1978). As always, provocative. *Journal of Educational Measurement, 15,* 297-300.

Raymond, M. R. & Reid, J. B. (2001). Who made thee judge? Selecting and training participants for standard setting. In G. J Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 119-157). Mahwah, NJ: Lawrence Erlbaum.

Shepard, L. A. (1979). Setting standards. In M. A. Bunda & R Sanders (Eds.), *Practices and problems in competency-based measurement* (pp. 72-88). Washington, DC: National Council on Measurement in Education.

Shepard, L. A. (1980). Technical issues in minimum competency testing. In D. C. Berliner (Ed.), *Review of research in education: Vol. 8* (pp. 30-82). Washington, DC: American Educational Research Association.

Shepard, L. A. (1984). Setting performance standards. In R. A. Berk (Ed.), *Criterion-referenced testing: Motives, models, measures and consequences* (pp. 169-198). Baltimore, MD: Johns Hopkins University Press.

Shepard, L., Glaser, R., Linn, R. & Bohrnstedt, G. (1993). *Setting performance standards for student achievement.* Stanford, CA: National Academy of Education.

Zieky, M. J. (1995). A historical perspective on setting standards. In *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments* n Statistics. (pp. 1-38). Washington, DC: National Assessment Governing Board and National Center for Education Statistics.

Zieky, M. J. (2001). So much has changed: How the setting of cutscores has evolved since the 1980s. In G. J Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 19-51). Mahwah, NJ: Lawrence Erlbaum.

About the Author

Robert L. Linn
School of Education
University of Colorado at Boulder
Email: Robert.linn@colorado.edu

Robert L. Linn is Distinguished Professor of Education at the University of Colorado at Boulder and Co-

Director of the National Center for Research on Evaluation, Standards, and Student Testing. He has published over 200 journal articles and chapters in books dealing with a wide range of theoretical and applied issues in educational measurement and has received several awards for his contributions to the field, including the ETS Award for Distinguished Service to Measurement, the E.L Thorndike Award, the E.F. Lindquist Award, the National Council on Measurement in Education Career Award, and the American Educational Research Association Award for Distinguished Contributions to Educational Research. He is past president of the American Educational Research Association, past president of the National Council on Measurement in Education, past president of the Evaluation and Measurement Division of the American Psychological Association, and past vice-president for the Research and Measurement Division of the American Educational Research Association. He is a member of the National Academy of Education, a Lifetime National Associate of the National Academies, and serves on two Boards of the National Academy of Sciences.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State Univeristy

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

This article has been retrieved 1367 times since September 5, 2003

Volume 11 Number 32 **September 5, 2003** **ISSN 1068-2341**

Charter Schools and Race: A Lost Opportunity for Integrated Education

Erica Frankenberg
Harvard University

Chungmei Lee
Harvard University

Citation: Frankenberg, E. and Lee, C. (2003, September 5). Charter schools and race: A lost opportunity for integrated education. *Education Policy Analysis Archives, 11*(32). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n32/.

Abstract

Segregation patterns in the nation's charter schools are studied. After reviewing state charter legislation that directly addresses issues of racial and ethnic balance of student enrollment, we briefly examine the racial composition and segregation of the charter school population nationally. School-level analyses, aggregated by state constitute the primary method of studying segregation in charter schools. First, we look at racial composition and segregation of charter schools by state. Then, we consider the differences in segregation between non-charter public schools (or simply "public schools" for convenience) and charter schools, as well as segregation within the charter school sector. We conclude with a discussion of the article's findings and recommendations to promote further racial equity in this growing sector of public schools. (Note 1)

Foreword

By Gary Orfield
Harvard University

Charter schools are one of the most important educational innovations of this generation. They have spread rapidly across the country and are often supported with fervent assurances that they can solve problems attributed to school bureaucracies. They are usually small, deregulated, run, at first, by a founder with a vision or a private company, and with faculties that are not supposed to be afflicted with the burnout and cynicism found in some high poverty schools with aging teachers. (Note 2) Embraced by both political parties, funded from federal, state, and local budgets, approved by most state legislatures, featured in countless newspaper stories, hailed as the potential antidote to all that is pathological in weak public schools, charter schools were put forward as something that combined the independence and autonomy of the private schools with public support and free tuition of the public schools. Many communities have been willing to try the experiment. According to the National Center for Education Statistics, there were 2,348 charter schools during the 2001-02 school year. (Note 3) Although there was an early concern that charter schools would serve as a haven for white students to escape diverse public schools, many minority parents have also expressed strong interest in alternatives to their local public schools and some minority led civil rights organizations run charter schools. (Note 4)

This article looks at only one aspect of the charter school story—whether or not these schools offer a less segregated experience than the public schools to the increasing numbers of students they serve. Obviously, this is but one of a number of dimensions on which these schools should be examined. Public schools have struggled with the issue of racial segregation for the past 50 years. We are now 15 years into an era of resegregation of our nation's schools, and black and Latino students are more isolated than they have been for three decades. This increasing isolation is not just isolation by race but also by poverty and, increasingly for Latinos and some Asian groups, by language. As reported in our latest study on national segregation trends, nearly nine-tenths of intensely segregated black and Latino schools have student bodies with concentrated poverty. (Note 5) The inequalities inherent in schools that serve children with worse health care, weaker nutrition, less educated parents, more frequent moves, weaker preschool skills, and often more non-English speakers are exacerbated by the fact that these schools are also less likely to have credentialed and experienced teachers. Since there is a very strong general relationship between segregation by race and poverty and educational inequality on many dimensions, this isolation can have serious consequences for students.

This article details a disappointing set of findings regarding its central question— charter schools are largely more segregated than public schools. Segregation is worse for African American than for Latino students, but is very high for both. In some states, white student isolation in charter schools is as high as that of African Americans. The problems reported here may not be due either to the intent or the desires and values of charter school leaders. They may reflect flaws in state policies, in enforcement, in methods of approving schools for charters, or the location where charter schools are set up.

The justification for segregated schools as places of opportunity is basically a "separate but equal" justification, an argument that there is something about the schools that can and does overcome the normal pattern of educational inequality that afflicts many of these schools. Charter school advocates continually assert such advantages and often point to the strong demand for the schools by minority parents in minority communities, including schools that are designed specifically to serve a minority population. It is certainly true that minority parents are actively seeking alternatives to segregated, concentrated poverty, and low-achieving public schools. (Note 6) White parents have also shown strong interest in educational alternatives as evidenced by the strong demand for magnet schools.

Unfortunately, despite claims by charter advocates, there is no systematic research or data that show that charter schools perform better than public schools. Since charter schools embody wildly different educational approaches and since charter and public schools obtain their enrollment in very different ways, evaluation and comparisons between the two require very careful analysis. At a minimum, it is certainly safe to say that there is little convincing evidence for the superiority of charter schools over public schools in the same areas. In fact, some of the studies suggest that charter schools are, on average, even weaker. (Note 7)

Authorization of charter schools is different in each state that has approved them. Charters permit and even welcome an enormous variety of innovative educational approaches, though they support very traditional approaches as well. Some of the charter founders are idealistic education leaders with a great new idea, strong imagination and inexhaustible energy, while some are committed community activists who have longed to run their own schools, or to serve only one group in a community, and many are managed by corporations that hope to profit from their operation. For many charter school founders, there is an implicit assumption that less government control and oversight will produce positive educational benefits.

One of the problems in evaluating the academic effectiveness of charter schools is that their effect is normally examined by comparing them to regular public schools, but their student body and parent groups are not the same, which makes the comparison of academic achievement inaccurate. Even if one were able to control for income, parent education, and other relevant, easily measurable family resources, there are several kinds of selection bias that make such comparisons virtually impossible. First, the families who are informed enough to choose a school and make the effort to get their child to a more distant school every day are not the same as the families who do not. (Note 8) Second, charter schools commonly lack the expertise and programs to serve students who are English Language Learners or severely disadvantaged children such as those in Special Education. As these students tend to score lower on standardized tests, if students from lower achieving groups do not enroll, the school's average scores will tend to rise. (Note 9) Third, many charters seek applications from students they believe would succeed, or who would respond to their approach, while not recruiting others. Some schools have screening procedures that public schools are prohibited from using because the public schools are required to serve all students. These biases mean that even if there were higher test scores or lower dropout rates for charter schools it might well be because of selective recruitment—students from families with more resources and/or fewer students with special needs—than because of the school's superior educational approach.

Curiously, in an era in which tests and accountability have been the hallmark of education policy, there has been little serious accountability for charter schools. Theoretically charter schools must meet the terms of their charter or they will be terminated. In most states, however, there are few resources for oversight of schools and revocations of charters for educational failure, as opposed to financial problems, are rare. (Note 10) Often their impact on racial segregation is ignored by the policymakers, despite the growing body of research evidence that has documented a trend of segregation in charter schools. If there is no real evidence linking superior performance to educational program rather than admissions selectivity, looking at general characteristics of the student body that are usually linked to educational inequality, such as levels of segregation, certainly deserves attention. On this front, there is little positive to say about these schools.

One might well think that charter schools would have a better chance to be integrated than public schools. Like magnet schools a generation earlier, charter schools offer distinctive curricula and the opportunity to create and manage schools with freedom from many normal constraints in large districts. Unlike magnet schools, charter schools have the added advantages of even greater freedom to innovate and for the most part, are not tied to geographically fixed attendance boundaries in residentially segregated communities as are neighborhood public schools but can draw from wherever interested students can be found (in some places where school districts grant charters, they are limited to the school district boundaries.)

The high level of racial segregation in charter schools is not a surprise when viewed in light of segregation in many aspects of American life. Those who think that charter schools are inherently likely to be free of racial inequality need to reflect on the racial consequences of other markets operating in areas of housing, employment, health care, etc. , where the markets have worked more to perpetuate and spread racial inequality than to cure it. One could accurately say that the normal outcome of markets when applied to a racially stratified society is a perpetuation of racial stratification. This is why early educational choice programs were often found to produce white flight from integrated schools and to contribute to segregation in many school desegregation trials. (Note 11) Those experiences were apparently unknown or overlooked by designers and supporters of many charter school policies.

In looking at the data presented here it is worth considering the experience of magnet schools. There have been a handful of highly selective schools in American public school systems, such as Boston Latin, San Francisco's Lowell High, and New York City's Stuyvestant High, which have produced remarkable students for generations. Overall, however, choice of schools and specialized curriculum for schools (except for vocational schools) were very rare in the U. S. until desegregation policies produced the magnet school movement in the mid-1970s. Magnet schools, like charter schools, grew rapidly in response to federal grant programs. The magnet school programs funded by the Emergency School Aid Act, however, had desegregation policies while the federal charter school law did not. The charter school law was a movement backward to the unregulated choice policies common 40 years ago across the South and in many big cities. Those did not work to produce integration and charter school policies do not either.

Racial segregation in charter schools needs to be considered as both a critical problem and a lost opportunity. Experience shows, that segregation is not inevitable and that it is possible to produce quite different outcomes with appropriate civil rights policies. As we approach the 50th anniversary of *Brown v. Board of Education*, this issue should be addressed and resolved. If we are to be serious about the impact of charters on minority opportunity in American schools, we need to look with considerable suspicion on unfounded claims of sweeping benefits, insist that accountability be extended fully to this sector, and not reach conclusions on the basis of assumptions rather than evidence.

This article should broaden the discussion of the future development of charter schools. Certainly any publicly funded schools should not be run in ways that either intensify racial isolation or undermine integrated schools in integrated neighborhoods. Charter schools offer opportunities, like good magnet schools, to create successful and voluntary diversity. Clearly there are some very ambitious and attractive schools being created under these policies. But too many are separate and unequal. We hope that this article will stir discussion and action to help develop the positive aspects of this innovation and to build into the charter school movement a commitment to offering school opportunities to all students that better reflect the diversity in our society as well as the demands of colleges and workplaces where they must eventually succeed.

Introduction

In the school year 2000-01, 1,855 charter schools were operating in 34 states that had passed legislation authorizing the creation of charter schools. (Note 12) Charter schools educate fewer than one percent of all public school students yet can have a substantial local impact on surrounding districts in terms of student enrollment. Most of these charter schools are concentrated in a few states, and in most states are located in urban areas. Charter schools in the sixteen states covered in this article (see Table 6 for list of states and their enrollments) make up more than 95% of the population of charter school students. (Note 13) Among different states there is great variation in the percentage of minority students attending charter schools. (Note 14) One reason for this variation could be that charter school reform has been supported by a diverse array of politicians and educators. Nonetheless, as publicly-funded schools of choice, it is important to examine whether these schools offer white and minority students interracial exposure when segregation across the country is increasing for black and Latino students, and white students are more racially isolated than students of any other racial/ethnic group (Frankenberg, Lee & Orfield, 2003).

In the past, most educational choice options (such as magnet schools) arose from desegregation plans (American Institutes for Research, 1993). In I973 the U. S. Supreme Court extended desegregation requirements to northern and western cities. However, just a year later, the Court rejected the lower Detroit court's proposition that integrating minority students in heavily minority and rapidly changing districts required including the suburbs to produce long lasting desegregation. Big cities looking at demographic facts and seeing the conflict over mandatory reassignments of students in cities such as Boston looked for a way to accomplish desegregation through voluntary choice. The problem was that very few whites had ever voluntarily chosen to attend black schools or to transfer for integration purposes. The idea of the magnet schools movement was to create specialized schools that could offer unique opportunities that would create a demand for voluntary transfers from both white and minority students and result in a student population that would meet desegregation standards (American Institutes for Research, 1993). By establishing special programs and curricular offerings in inner-city areas, school systems used magnet schools and programs to attract white students to predominantly minority schools. This movement became central to the desegregation strategies of cities such as Milwaukee, Cincinnati, and Buffalo. Furthermore, a title was written into federal law offering funds for such schools when they served desegregation purposes. The idea led to the creation of many highly popular and often well-integrated schools in districts that had few such opportunities and was strongly supported by school superintendents and boards. Because of the explicit emphasis on racial/ethnic balance, magnet schools are often among the most integrated schools in their district (Blank, Levine and Steele, 1996). By 2001 there were a reported 1,736 magnet schools in the county and there had been federal support for them for a quarter century. They enrolled 3. 0% of American students, compared to the 1. 2% in charter schools (Hoffman, 2003). Thus, magnet schools were a well-established model long before the charter school movement began.

Recently educational choice options have proliferated, through the growth of charter schools, vouchers, inter- and intra-district choice, magnet schools, and private schools. Building on the increasing belief of the importance of parents to have choice in their child's education, the No Child Left Behind Act (NCLB) of 2001 (Public Law 107-110) further expanded school choice by allowing students in failing schools to transfer. (Note 15) The rationale is that the achievement of poor and minority students will improve if they have access to schools that have demonstrated higher levels of student performance. It also assumes that parents will be able to make decisions about what education is best for their children, which will force schools to compete—and ultimately improve—to keep and/or attract students.

Charter schools, a form of school choice that is almost a decade old, represent a further attempt to institute school choice within the public education sphere. A charter is merely a political, legal, administrative and financial arrangement of relative autonomy, created in a somewhat different form in each state that has authorized them. The belief is that by introducing such choice options into the public schools, students and their parents could choose the school that was most appropriate, which would create incentives for all schools to improve in order to compete for students (Apple, 2001). Literature on school choice is mixed as to whether these assumptions are correct and actually result in improved education for all (for a brief discussion see Kim and Sunderman, 2003).

Since its inception, the charter school movement has been politically charged for both its proponents and opponents. Even within the charter school reform there is a diffuse group of supporters who favor charter schools for widely varying reasons; two of the major driving forces behind the charter school reform have been the excellence movement including high standards for all students and market-driven reforms aimed at making schools more efficient. Charter supporters say that such schools give important new options for parents, allow for educational innovation, and are not constrained by typical school district boundaries and student assignment practices that produce segregated patterns of schooling in many neighborhood school systems (Finn, Manno & Vanourek, 2000; RPP International, 2000).

Politicians have also supported the rapid growth of charter schools: NCLB also provided monetary assistance to increase the number of charter schools in states with charter legislation. (Note 16) In fact, while many public schools and districts across the nation are facing substantial budget cuts, President Bush has proposed $700 million in spending for charter schools.

Critics of school choice, however, argue that competition among schools will only improve student achievement if all schools are able to compete and students are equally free to choose. Otherwise, those students who are left behind by those who choose or are chosen in more competitive environments will have even less resources with which to compete (Arsen, Plank & Sykes, 1999). Those opposed also suggest that choice systems can compromise the public good by educating students in isolation from others for their private good, often further stratifying students along racial and socioeconomic lines (Cobb & Glass, 1999; Horn & Miron, 2000; Hochschild & Scovronick 2003). Additionally, school choice policies can allow schools to exclude students with special needs if it does not fit within their mission (Howe and Welner, 2002). The theory of market solutions rests on assumptions about choice in charters— that there is full information for everyone, that there are not economic or other barriers to attendance, an that the school will welcome students from all backgrounds. A great deal of experience with choice plans and magnet schools before the charter experiment show that knowledge and access were often very unequal, that families with the most resources and information often got access to the most highly regarded schools, that students from other races often felt unwelcome unless there were special efforts to recruit and support them in the new school, and that free transportation was essential to assure choice and access for lower income families (Fuller and Elmore, 1996). Choice plans that operate without these kinds of supports permit choice for only those who know what the choices are, how to access them, and do not need support to get to school.

Given these claims and counterclaims about charter schools, there remain important questions that should be addressed, particularly in the area of racial/ethnic segregation of students, which has been largely ignored in the overall debate about charter schools. Are charter schools offering students better opportunities for interracial exposure than the increasingly segregated public schools?Past research has shown that minority students attending integrated schools are more likely to attend and succeed in college, as well as to live and work in interracial settings (Wells & Crain, 1994; Eaton, 2001; Braddock II, 1980). Additionally, recent research by The Civil Rights Project has documented a number of important educational and civic benefits for students of *all* races in desegregated high schools (Kurlaender & Yun, 2001).

Segregated minority schools, where minority students experience little interracial exposure, are highly correlated with schools of concentrated poverty. Eighty-six percent of the students in all public schools that have greater than 90 percent black and Latino students of their total enrollment are also in schools where at least half of the student body is poor. (Note 17) These schools are more likely to have lower average test scores, less qualified and experienced teachers, and fewer advanced courses (Young & Smith, 1997). Moreover, research on charter schools has shown that charter schools with higher proportions of minority students tend to have fewer resources and less academic curricula than charter schools serving mainly white students (Fuller, Gawlik, Gonzales, Park & Gibbings, 2003). As publicly-funded schools, it is essential that charter schools provide equal educational opportunity for all students.

Charter school proponents claim that charter schools provide options for low socio-economic students (Finn, Manno & Vanourek, 2000). Preliminary analyses question whether charter schools are, in fact, achieving this goal of educating low-income students. At the national level, in 1997-8, 39% of charter school students versus 37% of public school students received free and/or reduced lunch. Miron and Nelson (2002) report that, based on data from half of Michigan charter schools, when examining the student poverty composition of charter schools in comparison to their surrounding districts, charter schools serve a slightly lower percentage of low-income students; there are similar and even stronger trends in California, Massachusetts, and Colorado (SRI International, 1997; Wood, 1999; Clayton Foundation, 1999). At the district level, Ascher and colleagues (1999) found that only 35% of charter schools were socio-economically diverse (between 20% and 80% of students on free/reduced lunch) as compared to 72% of public schools in surrounding districts. However, it is difficult to determine the level of student poverty in charter schools because many of these schools do not participate in the federal free/reduced lunch program, which is the most common measurement of the socio-economic status of students (Wells, Holme, Lopez, & Cooper, 2000). (Note 18) More analysis is needed to accurately ascertain the levels of student poverty in all charter schools—particularly in the many segregated charter schools that exist across this country.

Research Questions

Recently, issues of accountability and equity for charter schools have come under greater scrutiny (Cobb & Glass, 1999; Wells, 2002). However, as the 2001 RAND book, *Rhetoric versus Reality: What We Know and Need to Know about Vouchers and Charter Schools*, concludes, given the different conditions under which charter schools operate, we really do not know much about the issue of racial segregation in charter schools (Gill, Timpane, Ross, & Brewer, 2001). Because of their complexities—they are both public but also independent from the public school system, they can choose their students but also need to attract students, and they are governed by state charter legislation yet are influenced by their local context and mission—it is difficult to know how to even evaluate charter schools.

There has been some research to suggest that black students have a relatively high level of access to charter schools (Wells, et al. , 2000). Recent state evaluations have found that even though the aggregate racial composition of charter schools is similar to host districts, there are great differences at the school level in enrollment compositions (Cobb & Glass, 1999; Miron, Nelson and Risley, 2002).

In this article, we address one key aspect of the multi-faceted charter school phenomenon with the following questions:

- What is the racial/ethnic composition of charter schools?
- What is the average exposure of charter school students to students of other races in their schools?
- How are charter school students distributed among the charter schools?
- Are students more racially isolated in charter schools than in public schools?

There is strong evidence that many Americans believe in the importance of integrated education. Sixty percent of blacks in 1998 and 34% of whites believed that it is "absolutely essential" for schools to "have a diverse student body with kids from different ethnic and racial backgrounds" (Farkas and Johnson, 1998). Further, a national poll in 1999 reveals that 68% of all respondents believe that integration had "improved the quality of education" for blacks and 50% believe that it had made education better for whites. By 1999, almost three-fifths of Americans believed that we needed to do more to integrate schools (Gallup, 1999). Certainly there is also substantial support for choice policies; in 1993, 65% of the public were in favor of allowing students and parents to choose what schools they attended, regardless of where they lived (Elam, Rose & Gallup, 1993). However, as subsequent discussion will illuminate, despite many parents' preferences for integrated schools and choice policies, many state charter laws are not explicitly supportive of racial diversity in charter schools.

Data and Methods

We compare the racial composition of charter schools with that of all non-charter public schools by examining who is enrolled in charter schools and the extent to which they are segregated. Although in 2000-01 charter schools enrolled fewer than one percent of all public school students in the country, many of these schools are concentrated in certain areas and states, and can have a substantial local impact on surrounding public school district enrollment and racial diversity. We focus on the sixteen states that had total statewide charter enrollments of at least 5,000 students in 2000-01. Charter students in these sixteen states account for 95. 4% of the entire U. S. charter school population. The data analyzed for this article are from the National Center for Education Statistics 2000-01 Common Core of Data (CCD). The CCD is a comprehensive, yearly national dataset of all operational public schools and includes school information on student characteristics such as enrollment and racial counts (Note 19) that are comparable across states and between charter schools (Note 20) and non-charter public schools.

In examining issues of charter school segregation, we use several measures to evaluate different school-level dimensions of segregation. By aggregating the school-level data to the state level we are able to compare charter and public schools within a particular state as well as charter school segregation across states. The exposure index provides an average picture of the interracial exposure of students: the index can be interpreted as the percentage of students of a particular racial group in the school of the average student of another group (Massey & Denton, 1988; Orfield, Bachmeier, James & Eitle, 1997; Reardon & Yun, 2002). For example, Michigan's charter school white-black and white-white exposure rates (Note 21) of 16% and 78%, respectively, (Table 8), mean that, on average, Michigan's white charter school students attend a school where 16% of students are black and 78% of the students are white. If students were evenly distributed (e.g., no black-white segregation), all Michigan charter school students would, on average, attend schools that are 54% black and 40% white, respectively, a racial composition equal to the proportion of white and black students in Michigan's total charter school enrollment (Table 6). These exposure indices demonstrate that white charter school students in Michigan, on average, attend schools that disproportionately enroll high levels of white students and low levels of black students.

Examining the exposure index gives us an average picture of interracial exposure in charter schools. However, this measure, which is essentially a weighted average of the racial composition of schools of students from each race, can mask the variation and distribution of students in schools. For example, if black exposure to white students in charter schools is 50%, that could describe two schools that are both 50% white, or could be one school that is 90% white and one school that is 10% white. These two examples would have very different implications in terms of the interracial experience of students in charter schools. To explore the distribution of students in charter school, we examine the concentration of students of all races in predominantly minority schools (greater than 50% of the student body is non-white), intensely segregated minority schools (90-100% minority), and intensely segregated white schools (90-100% white). Together, these measures portray both the actual level of interracial exposure i schools as well as the percentage of students attending racially imbalanced and isolated schools.

It is important to note that using schools as our unit of analysis, this article analyzes the racial composition and exposure at the state level. Previous studies at the district- and school- level have shown that when examined in terms of their local contexts (comparing the racial enrollments of charter schools to that of the surrounding public school district or the closest public school), charter schools are less racially diverse than local public schools and districts (Wells, et. al, 2000; Ascher, Jacobwitz, & McBride, 1999; Cobb & Glass, 1999). We recognize that the context of where schools are situated locally and how districts choose to interpret state charter legislation are important considerations that likely influence the outcomes we examine. However, we do not specifically address that in this article. One reason is because our data do not allow us to examine these questions at a more local level. However, it is potentially misleading to look at charter schools at the district level, (Note 22) because in many states charters are not necessarily part of a school district or confined to drawing students only from surrounding districts. (Note 23)

One characteristic common across all charter schools is the statewide nature of charter school legislation. This orientation influences the context in which all charter schools throughout the state must operate. In addition, who can attend charter schools, how many can be established, and by what means they enroll students are just some of the stipulations in charter school legislation that differ widely among states. Demographic contexts of the entire state population also vary across the country and these variations can affect the racial composition of the students in charter schools. Furthermore, although charter schools can enroll students across district and county lines throughout metropolitan areas, charter schools do not enroll students across state lines. A comparison between charter schools and public schools at the state level gives us important comparisons of the racial composition and segregation in the small but growing sector of charter schools within legislatively defined geographic boundaries Our purpose in this article is not to discount the variation that occurs at the district- and school-level, but simply to focus on state-level observations of differences in racial composition between public schools and charter schools and how students are distributed among charter schools.

Findings

In the sixteen states with charter school enrollments greater than 5,000, we find that charter schools in most of these states enroll disproportionately high percentages of minority students, particularly African American students. Over half (56%) of all charter schools in these states are located in central cities. Specifically, we find the following trends for charter school students by race:

- Seventy percent of all black charter school students attend intensely segregated minority schools compared with 34% of black public school students. In almost every state studied, the average black charter school student attends school with a higher percentage of black students and a lower percentage of white students.
- White students in every state studied attend schools with a much higher white percentage than their overall share of the charter school population. In many states, however, white charter school students are exposed to substantial percentages of non-white students. Furthermore, there are pockets of white segregation where white charter school students are as isolated as black charter school students.
- The pattern for Latino segregation is mixed; on the whole, Latino charter school students are less segregated than their black counterparts.

In sum, although many of the charter laws require compliance with desegregation orders or mandate specific racial/ethnic balance in charter schools, there is little evidence of serious effort at the state level to ensure racial balance.

Charter Legislation

In a reform with such a diverse array of schools and ideologies, one of the few consistencies for the charter schools in a state are the state charter school legislation and guidelines under which all schools are supposed to operate. This legislation and regulations vary significantly among states. More than half of all states with charter school laws have policies that require charter schools to comply with desegregation standards or reflect student racial/ethnic populations in the state (see Table 1). In most cases, the state or local education agency (usually a school district or the state department of education but can vary in some states), and not the state itself, authorizes the charter schools and reviews and regulates the schools.

Although the charter school reform is primarily governed by policies set by each state, there are federal regulations and programs that may also affect the composition of the student body of charter schools. In 1994, a new federal grant program was implemented to support charter schools as part of the Improving America's Schools Act. (Note 24) Charter schools can receive funding through federal programs such as, the Eisenhower Professional Development Program, the Safe and Drug Free Schools Act, and the Perkins Occupational Education Act. However, federal funding can only be used if charter schools comply with federal civil rights statutes such as Title VI. NCLB provides funding to schools with high levels of student poverty (formerly known as Title I of the Elementary and Secondary Education Act) —but accepting NCLB money means that these schools must comply with federal civil rights provisions. (Note 25) Likewise, althoug states individually pass their own charter legislation, if charter schools receive money from the federal Public Charter Schools Program, they are required to use a lottery to admit students in the event that there are more applicants than available slots for the school.

Courts have held that in school districts under federal court desegregation orders, charter schools will not be allowed to impede compliance with a court's or administrative entity's desegregation plan. (Note 26) However, even in this instance, the federal guidelines as to the responsibilities of charter schools in such situations are unclear, at best. (Note 27) Moreover, there is not a general framework to support such efforts. Charter schools are often given little support in implementing these guidelines, and in general, there is not a framework to support those who value racially diverse schools (Wells, 2002). (Note 28)

Of all states with charter school legislation, nineteen states have specific racial/ethnic balance enrollment guidelines for their charter schools (Table 1). (Note 29) Without these rules, charters have little incentive to maintain racial/ethnic balance in their schools. Two of the four states with the largest enrollment of charter school students (Arizona and Texas) have no racial/ethnic guidelines. There are also some states that include equity provisions—such as providing free transportation to all students or requiring information to be widely available—that are important in ensuring that students from all backgrounds are truly able to choose to enroll in charter schools. Nine states with racial balance policies are included in the state-level analysis of this article (those states with charter enrollment greater than 5,000). Interestingly, six of the nine states in our analysis that have specific racial/ethnic guidelines are southern states. (Note 30) Among the seven southern states in our group of sixteen states with at least 5,000 charter students, only Georgia has no racial balance provision.

Moreover, the language of the racial/ethnic balance provisions varies from state to state. In some states, general guidelines regarding non-discrimination on the basis of race is used; fewer than ten states require compliance with desegregation orders. We find that despite the specific racial/ethnic balance guidelines in charter legislation, many states still have racially imbalanced enrollments. Because many state regulations call for district proportionality and this analysis is primarily state-level, more research is needed at the district level to determine the impact of the guidelines. Perhaps even state charter laws with racial/ethnic balance language are still too weak; without other equity provisions built in to this market-based reform, charter schools are unlikely to overcome the persistent segregation of our larger society.

Table 1
Racial/Ethnic Guidelines in State Charter School Legislation, 2003(Enrollment 2000-01)

State	Enrollment	Charter Legislation
Alaska	2,594	State law contains no discrimination provisions other than general non-discrimination provision.
Arizona	45,596	State law contains no discrimination provisions other than general non-discrimination provision.
Arkansas	708	Charters in districts under court-ordered desegregation plans must use a weighted lottery in student selection as well as issues relative to funding.
California	112,065	Charter must specify means by which a school's student body will reflect racial and ethnic balance of the general population living in the school district granting the charter.
Colorado	20,155	A charter school shall be subject to any court-ordered desegregation plan in effect for the school district in which it operates.
Connecticut	2,429	Charter must specify procedures to promote a diverse student body and state board will give preference to granting charters in districts that have 75% or more minority students.
Delaware	2,716	Charter school may not be formed to circumvent a court-ordered desegregation plan.
District of Columbia	**	State law contains no discrimination provisions other than general non-discrimination provision.
Florida	26,893	Racial/ethnic balance of charter school may not differ from district or community.
Georgia	20,066	State law contains no discrimination provisions other than general non-discrimination provision.
Hawaii	1,343	Charter must include plan for identifying, recruiting, and selecting students to make certain that student participation is not exclusive, elitist, or segregative.
Idaho	1,083	State law contains no discrimination provisions other than general non-discrimination provision.
Illinois	7,552	State law contains no discrimination provisions other than general non-discrimination provision.
Indiana	0	Charter school must have plan for compliance with any applicable desegregation order.
Iowa	0	State law contains no discrimination provisions other than general non-discrimination provision.
Kansas	67	Pupils in attendance at the school must be reasonably reflective of the racial and socio-economic composition of the school district as a whole.
Louisiana	3,212	Must comply with any desegregation order/regulation.
Massachusetts	13,712	State law contains no discrimination provisions other than general non-discrimination provision.
Michigan	54,751	Must comply with any desegregation requirements.
Minnesota	9,395	If the charter school reflects the racial and ethnic diversity of the area, it may limit admission to a geographic area of greater than average non-white population.
Mississippi	367	State law contains no discrimination provisions other than general non-discrimination provision.
Missouri	7,061	Admit district residents provided that such preferences do not result in the establishment of racially or socio-economically isolated schools.
Nevada	1,255	Racial balance of charter school may not differ from district by more than 10%.
New Hampshire	0	State law contains no discrimination provisions other than general non-discrimination provision.
New Jersey	10,179	Charter must have a plan to enroll cross-section of school-aged population including racial and academic factors. Commissioner of Education must assess whether charter will have segregative effect

		on district of residence of the charter school, and after the charter is operating, Commissioner must assess whether charter has a segregative effect on other districts sending pupils to the charter.
New Mexico	1,335	State law contains no discrimination provisions other than general non-discrimination provision.
New York	***	State law contains no discrimination provisions other than general non-discrimination provision.
North Carolina	15,523	After one year, charter school must reasonably reflect racial balance of district, and the school will be subject to any court-ordered desegregation plan in effect for the school district in which it operates.
Ohio	14,745	Community school shall achieve racial and ethnic balancereflective of the community it serves.
Oklahoma	1,208	Charter school may not admit student who resides in school district under court desegregation order or relevant US Department of Education OCR agreement if resident school district notifies charter school that admission of said student would violate order or agreement.
Oregon	559	State law contains no discrimination provisions other than general non-discrimination provision.
Pennsylvania	18,981	School district may not approve charter application if charter school would place the school district out of compliance with a desegregation order of a federal or state court order or a state human relations commission order.
Rhode Island	557	Charter school must have a program to encourage the enrollment of a diverse student population, and the makeup of the school must be reflective of the population of the district.
South Carolina	483	Racial composition of charter school enrollment may differ by no more than twenty percent from school district or targeted student population, but local school district may find charter school not operating in racially discriminatory manner without regard to twenty percent requirement.
Tennessee	0	State law contains no discrimination provisions other than general non-discrimination provision.
Texas	37,978	State law contains no discrimination provisions other than general non-discrimination provision.
Utah	537	State law contains no discrimination provisions other than general non-discrimination provision.
Virginia	55	Charter must comply with any desegregation orders/regulations.
Wisconsin	9,511	Racial balance of charter school may not differ from district.
Wyoming	0	Racial balance of charter school may not differ from district, and the means by which this balance is to be achieved must be specified in charter.

Note: Some states as of 2000-01 had passed charter school legislation but there were no charter schools yet operational. Thus, some states (e. g. , New Hampshire) have no enrollment as of 2000-01. Source: Statutes concerning charter schools are found using Westlaw and Lexis-Nexis. The specific citations are available upon request from the authors.

** In 2000-01 District of Columbia had 33 charter schools, but did not report student data to NCES.

***In 2000-01, New York had 38 charter schools, but did not report student data to NCES.

National Trends (Note 31)

While they can mask considerable variation among the states' implementation of the charter school reform, national statistics provide a helpful background in which to consider charter school students and their distribution among schools. In the 34 states with charter schools in 2000-01, less than half (43%) of all charter schools students were white. Another one-third (33%) were black and one-fifth (19%) were Latino. Asian and Native American students make up a very small percentage of the charter school enrollment. The national non-charter public school population has a much higher percentage of white students (a difference of sixteen percentage points) and a lower percentage of black students than charter schools (Table 2). (Note 32) The percentage of black students in charter schools is almost twice the total black public school enrollment. The share of Latino students in charter schools versus public schools is comparable. The fraction of Asian students in charter schools is slightly less than their proportion of the total public school population, while that of Native Americans is slightly more.

Table 2
Enrollment and Racial Composition of Charter and Public Schools, 2000-01

	Enrollment	White(%)	Black(%)	Latino(%)	Asian(%)	Native American (%)
Charter	444,825	43	33	19	3	2
Public	36,116,860	59	17	19	4	1

Almost ninety percent of black charter school students are in predominantly minority schools where minority students are more than 50% of the student body (see Table 3). Seventy percent of all black charter school students, over 100,000 students, are in 90-100% minority charter schools. This number is striking when compared to the 34 percent of black public school students who attend 90-100 % minority schools. Although the public school figure (34%) is the highest it has been in three decades, the charter school distribution suggests even more segregation. (Note 33) These numbers indicate that black students are not only disproportionately over-enrolled in charter schools, but that they are enrolled at a much higher rate than other black public school students in intensely segregated minority schools.

White charter school students are also more likely to be in predominantly minority and intensely segregated minority schools than white public school students. The percentage of white students in such schools, however, is much lower than students of any other race, in both charter *and* public schools. (Note 34) Higher percentages of Latino and Asian charter school students attend intensely segregated minority schools than their public school peers, but their rates of attendance in predominantly minority schools are similar.

Table 3
Percentage of Charter and Public School Students in Segregated Minority Schools, by Race/Ethnicity, 2000-01

	Charter			Public	
	50-100% Minority	90-100% Minority		50-100% Minority	90-100% Minority
White	17	2		13	1
Black	89	70		71	34
Latino	78	42		77	37
Asian	57	21		56	14
Native American	65	45		47	19

Eighty-three percent of white charter school students are in majority white schools (Table 4). About one-fifth (22%) of all white charter school students nationwide are in schools that have a student body that is more than 90% white, a rather high percentage due to the fact that the majority of students in charter schools are minority students.

Not surprisingly, given their high concentration in minority schools, black charter school students are the least commonly found in predominantly and intensely segregated white schools. Ten percent of black charter school students attend majority white schools and only about one percent is in 90-100% white charter schools. These rates are substantially lower than those of students of other racial groups except Latinos. Interestingly, Latino students are the most segregated from whites in public schools, but Latino charter students—while still highly segregated from white students—are less segregated than black charter students. Just over one-fifth (22%) of all Latino charter school students are in majority white schools, twice the percentage of black students in such schools. While Native American public school students are exposed to a higher share of white students than students of any other minority group, in charter schools, Asian students are more commonly enrolled in white schools than other minority students.

Table 4
Percentage of Charter and Public School Students in Segregated White Schools, by Race/Ethnicity, 2000-01

	Charter			Public	
	50-100% White	90-100% White		50-100% White	90-100% White
White	83	22		88	39
Black	11	1		29	2
Latino	22	1		23	2
Asian	43	5		44	6
Native American	36	3		54	8

Nationally, the average white charter school student attends a school that is 72% white. White exposure to black and Latino students is fairly even: the percentage of black and Latino students in the average white charter student's school is 12 and 11 percent, respectively (see Table 5). White exposure to other racial minorities is low, in part due to the small percentages of Asian and Native American students attending charter schools.

On average, black and Hispanic students are disproportionately exposed to higher percentages of students of their own race in charter schools. For example, the average black charter school student attends a school that is 73% black and only 14 percent white. The percentages of Latino and white students in the charter school of the average Latino are 52% and 26%, respectively. Perhaps due to their low enrollment in charter schools, Asians and Native Americans are exposed to more whites than are either black or Latino students.

Table 5
Racial Composition of Schools of the Average Charter School Student, by Race/Ethnicity, 2000-01

Percent Race in Each Charter School	Racial Composition of Charter School Attended by Average:				

	White Student	Black Student	Latino Student	Asian Student	Native American Student
% White	72	15	26	43	32
% Black	11	73	18	16	7
% Latino	12	11	52	19	11
% Asian	3	1	3	20	2
% Native American	1	0. 4	1	1	48
Total	99	99	100	99	100

Note: Totals may not add to 100 due to rounding.

In sum, at the national level, blacks are over-enrolled and whites are under-enrolled in charter schools relative to public school enrollment. Black charter school students are overwhelmingly found in intensely segregated minority schools, and are more segregated from white students than black public school students. However, for white charter school students, the story is quite different. Because whites make up a relatively small percentage of the charter school population, they are exposed to more blacks and Latinos and to fewer white students in charter schools than in the public schools at the national level. (Note 35) For Latino students, at the national level, public and charter school segregation rates are similar.

Because aggregation of racial composition and segregation at the national level can obscure more localized variation, it is also important to see how these trends vary by state between charter and public schools. The over-enrollment of black students in charter schools indicates segregation between charter schools and public schools; therefore, it is important to also examine the distribution of students within the charter sector. This paper, then, looks at each of these issues in turn.

State-Level Trends: Racial Composition

In 2000-01 there were sixteen states with at least 5,000 students in charter schools, but the number of students enrolled in these schools and the racial composition of the schools varied widely across states. California, the most populous state, has the largest charter school population with over 100,000 students in charter schools during 2000-01. On the other end of the spectrum, there are 18 states whose charter school enrollment totals less than 5,000 and are not included in our state-level analysis. Of the sixteen states with substantial charter school enrollment, nine have guidelines specifying racial balance in the state charter school legislation (see Table 1 above for racial/ethnic balance guidelines in all states with charter legislation).

Among all public school students, only six states have a majority non-white student body (Frankenberg, Lee & Orfield, 2003). For charter school students, the picture is very different: only six of the states with a substantial charter population have a majority of the charter school enrollment that is *white* (see Table 6). In fact, six states have charter enrollments that are more than 50% black. Eight states have at least 15% of the charter school enrollment composed of Latino students. Asian students account for a very small percentage of students enrolled in charter schools; only in California and Minnesota are Asian enrollments greater than 5% of the total charter population. The Native American population is also small in all states except Minnesota and Arizona. (Note 36)

Table 6
Enrollment and Racial Composition of Charter Schools by States with more than 5,000 Charter School Students, 2000-01

State	State Total	White (%)	Black (%)	Latino (%)	Asian (%)	Native American (%)
Arizona	45,596	56	8	27	2	8
California	112,065	42	18	34	5	1
Colorado	20,155	74	7	16	2	1
Florida	26,893	50	31	18	1	0
Georgia	20,066	64	28	6	3	0
Illinois	7,552	9	68	23	1	0
Massachusetts	13,712	54	27	15	3	1
Michigan	54,751	40	54	4	1	1
Minnesota	9,395	52	23	5	15	6
Missouri	7,061	9	85	5	1	0
New Jersey	10,179	12	71	15	2	0
North Carolina	15,523	53	43	2	1	1

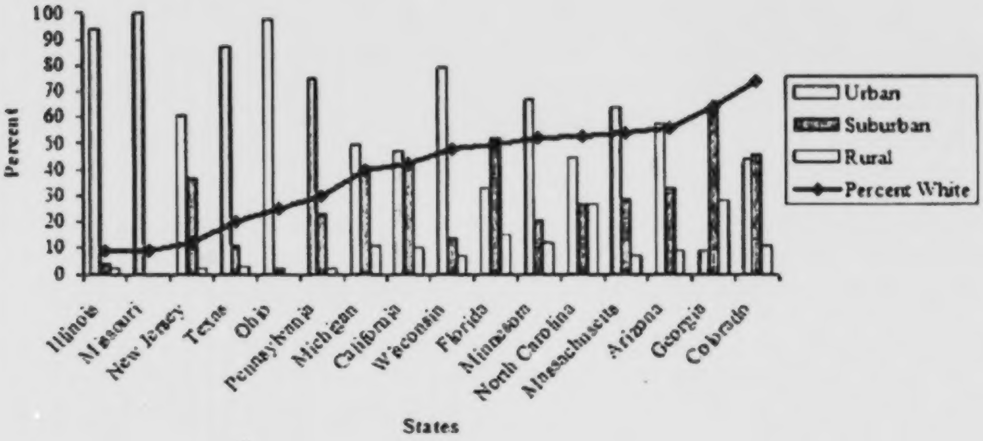
Ohio	14,745	25	73	1	0	0
Pennsylvania	18,981	30	61	8	1	0
Texas	37,978	20	41	37	1	0
Wisconsin	9,511	48	38	8	5	1

In the sixteen states included in this study, charter schools were predominantly located in cities. Table 7 displays the location of charter school students in each of the sixteen states, ranked by percentage of white charter students in each state. (Note 37) States with higher percentages of charter school students in cities were less likely to have large white enrollments, similar to trends in large central city public school districts. (Note 38) Overall more than half of the charter school students in these sixteen states attended schools that were located in central cities (56%) while a third (34%) were in schools located in suburban areas. Missouri (100%), Ohio (98%), Illinois (94%), and Texas (87%) had the highest proportion of their charter school students in cities and were four of the five states with the lowest percentage of white students of their total charter enrollment. Charter schools in these four states educated almost one-sixth of all charter school students. Only three states, Florida (52%), Georgia (63%), and Colorado (46%) had greater percentages of their charter school students enrolled in schools in suburbs than in cities. (Note 39) Generally, as can be seen from Figure 1, states with the lowest proportion of white students in their charter schools also had the highest proportions of their charter school students in central city schools while states with the highest proportion of white charter school students were those that have higher proportion of charter school students enrolled in suburban areas.

Table 7
Percentage of Charter Schools by Location and State, 2000-01 (Ranked by Percent White)

State	White (%)	Urban (%)	Suburban (%)	Rural (%)
Illinois	9	94	4	2
Missouri	9	100	0	0
New Jersey	12	61	37	2
Texas	20	87	11	3
Ohio	25	98	2	0
Pennsylvania	30	75	23	2
Michigan	40	50	39	11
California	42	47	43	10
Wisconsin	48	79	14	7
Florida	50	33	52	15
Minnesota	52	67	21	12
North Carolina	53	45	27	27
Massachusetts	54	64	29	7
Arizona	56	58	33	9
Georgia	64	9	63	28
Colorado	74	44	46	11
Percent of Total		56	34	10

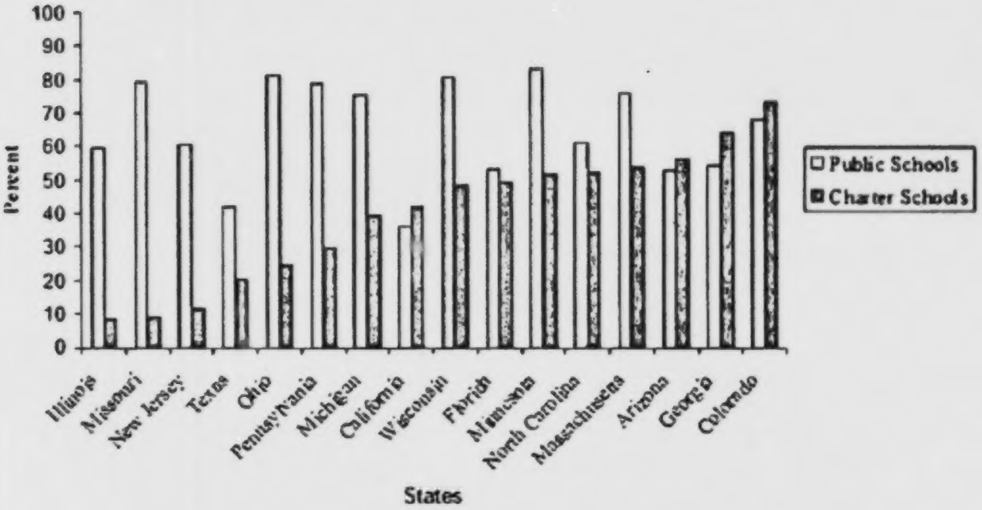
Figure 1: Enrollment of Students by Location and State, 2000-01



*Note:*States with zero percent of charter schools in a given location may have less than three bars. For example, 100% of Missouri's charter schools are in urban areas, so there is no suburban or rural bar for Missouri.

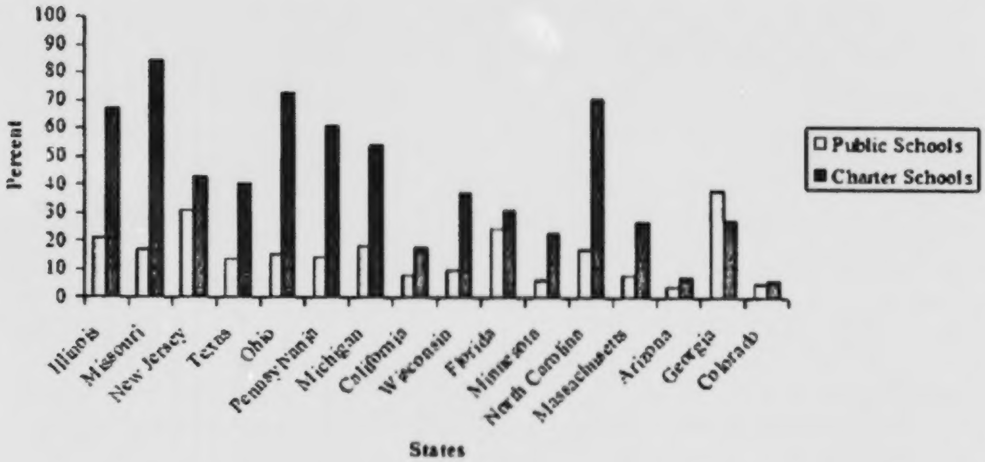
As mentioned above, the demographics of the states' populations and public school enrollments vary widely. Thus, we examine how the state's charter school racial composition compares to the state's public school enrollment by race. In almost every instance, the white percentage of charter school students is smaller than in public schools. In ten of these states, the white percentage in public schools is at least twenty percentage points higher than the white share of total enrollment in charter schools. Half of these states are Midwestern states. Missouri shows the starkest contrast between public and charter white enrollment: the white percentage of the public school enrollment is more than eight times greater than the white charter school proportion. (Note 40) Four states have a greater proportion of white students in charter schools than in public schools (see Figure 2).

Figure 2: White Proportion of All Students Enrolled in Charter Schools and Public Schools, 2000-01



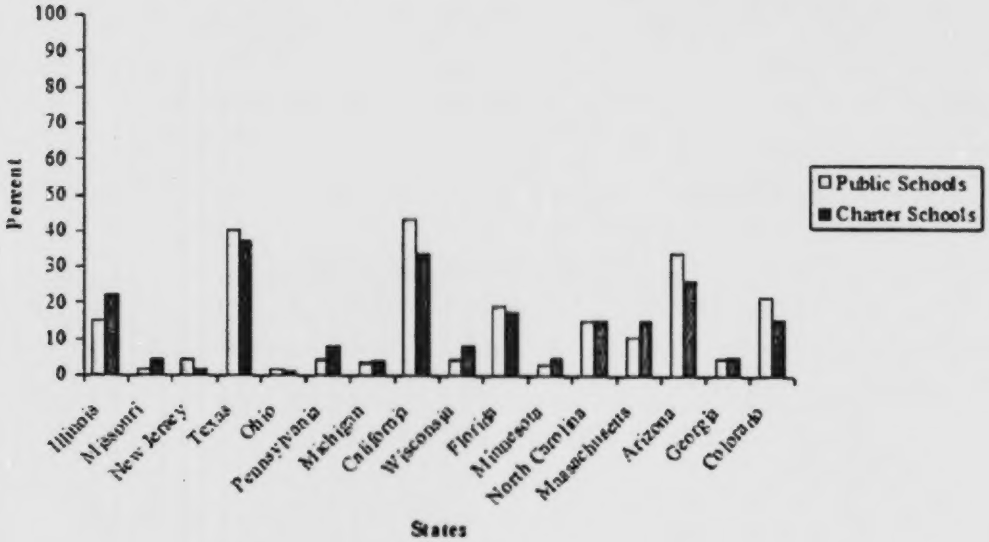
The reverse trend holds for black enrollment: in every state except Georgia, charter schools have a higher black enrollment share than public schools (see Figure 3). For example, in New Jersey, Ohio, and Missouri, although black students are less than 20% of total public school enrollment, black students make up more than 70% of charter students in these states, despite specific racial guidelines in the state charter legislation in all three states. Interestingly, Georgia has the highest black percentage of total public school enrollment and is the only state in which charter schools disproportionately enroll a lower proportion of black students.

Figure 3: Black Proportion of All Students Enrolled in Charter Schools and Public Schools, 2000-01



In most states, the differences between Latino public and charter school enrollment are far smaller than for white and black students (see Figure 4). The largest difference is in California, where the Latino portion of charter school enrollment (34%) is ten percentage points lower than the Latino portion of public school enrollment (44%). The states with the largest under-enrollment of Latino students in charter schools (California, Arizona, and Colorado) are all in the West.

Figure 4: Latino Proportion of All Students Enrolled in Charter Schools and Public Schools, 2000-01



Exposure to Students of Other Races

We have already documented that charter schools, when compared to public schools at the state level, disproportionately enroll higher percentages of black students and lower percentages of white students relative to non-charter public schools, which suggests that segregation between charter and public schools exists both nationally and state by state. It is critical to more closely examine these distributions, to see whether students are enrolled evenly across charter schools or whether they are isolated in schools with students of their own race. One commonly used measure of segregation is the exposure index, which describes racial composition of the school attended by the typical student of a given race.

White Student Exposure

Within charter school sector

As seen in Table 8, white students in every state attend schools with a much higher white percentage than their overall share of the charter school population. For example, although Missouri's white charter students are exposed to a lower percentage of white students (23% on average), this is more than twice the white share of charter enrollment (9%). Even in states where they are only a small percentage of charter school enrollment, whites are generally concentrated in schools with other white students and substantially isolated from students of other races. For example, in Illinois, Texas, and Ohio, where less than one in four charter school students is white, the average white charter student attends a school where more than 50% of the student body is white (Table 8). The isolation of white students in Illinois is particularly marked. Despite white students comprising less than 9% of the overall charter enrollment, the typical white student is in a school which is 54% white, a percentage that is six times higher than the white share of the state's charter school enrollment.

As a result of these relatively high levels of white isolation in charter schools, white students, in general, are exposed to lower percentages of students from other racial groups than would be expected by enrollment share alone. Except in four states with the highest black enrollment share (i. e. , Illinois, Missouri, New Jersey, and Ohio), the average white student attends a charter school where fewer than one in five students is black. Even in states where over half the charter school population is black, white students, on average, attend schools with more white students than black students. In Michigan, where over half (54%) of charter students are black, the average white student is exposed to five times as many white students as black students (white students, on average, attend charter schools that have 16% black students and 78% white students).

In most state, white exposure to Latino students in charter schools is lower than white exposure to black students, which might be due to the lower enrollment of Latino students in charter schools in some states. The four exceptions are in the West (Texas, California, Arizona, and Colorado). In ten states the average white charter school student attends a school with less than 10% Latino students. The high isolation of white charter school students in Illinois, however, prevents substantial white exposure to Latinos despite a relatively high charter school enrollment of Latinos. Illinois has the fourth highest percentage of Latino students in charter schools (23%); yet the average white student in Illinois attends a school that is only 9% Latino.

Charter vs. public schools

Regardless of the type of school (i. e. , charter or public), the average white student attends a school with a higher proportion of white students than the state's aggregate percentage of white students, which suggests some sort of segregation mechanism at work. (Note 41) However, as noted above, white charter students in ten states are less isolated than public school students. This could be due to a lower percentage of white students enrolled in charter schools than public schools in these states, which would make it more difficult to create schools that were predominantly white. However, in states where the white share of total enrollment is similar in both public and charter schools (Note 42) (California, Florida, North Carolina, Arizona, and Georgia), the average white charter student is equally as isolated or more isolated in schools with other white students than the average white public school student. This provides support to the contention that it is not that charter schools are inherently doing a better job of integrating students, but rather that low white enrollments are responsible for the lower levels of white racial isolation in charter schools in most states. (Note 43)

In terms of white students' exposure to minorities in charter versus public schools, white students in most states, on average, are more exposed to black students in charter schools than in public schools. In fact, the average white charter school student in all states except Colorado, Florida, North Carolina, and Georgia has greater exposure to black students (see Table 8) than does the average white public school student. This could be due to the disproportionately high enrollment of black students in charter schools.

There are few differences between public and charter schools for white exposure to Latino students. In four states—mainly in the West—white exposure to Latinos is lower in charter schools, than in public schools. However, these differences tend to be small.

Table 8
White Exposure in Public and Charter Schools, by State, 2000-01

States	Percent White		White Isolation		White Exposure to Blacks		White Exposure to Latinos	
	Charter	Public	Charter	Public	Charter	Public	Charter	Public
Illinois	9	60	54	82	34	7	9	7
Missouri	9	80	23	90	70	8	6	2
New Jersey	12	61	46	79	33	8	16	7
Texas	20	42	53	65	19	10	25	22
Ohio	25	81	64	91	33	7	2	1

Pennsylvania	30	79	75	90	19	6	4	2
Michigan	40	75	78	89	16	5	5	3
California	42	36	67	58	7	5	20	26
Wisconsin	48	81	76	89	13	4	7	3
Florida	50	54	71	69	15	17	12	12
Minnesota	52	83	83	89	7	4	3	3
North Carolina	53	61	79	71	18	22	1	4
Massachusetts	54	76	79	86	11	4	7	6
Arizona	56	53	74	70	5	4	16	21
Georgia	64	55	78	72	15	22	4	4
Colorado	74	68	81	77	4	4	12	15

Minority Student Isolation

Within the charter school sector

Black charter students are heavily isolated in overwhelmingly black schools. This could be due partially due to their disproportionately high enrollment in charter schools relative to non-charter public schools. However, black isolation indices are well above proportional representation (e. g. , black share of total enrollment), which suggest something in the structure of charter school enrollment that acts to segregate black students, such as the large percentage of charter schools located in central cities. The exposure and isolation indices for black and Latino students in charter and public schools are presented in Table 9. Except in two states (Arizona and Colorado), black charter school students, on average, attend majority black charter schools. In almost half of the states, the average black charter student attends a school that is at least three-quarters black. Illinois provides an interesting example. In Illinois, 68% of the charter school enrollment is black and the typical black charter student's school is 77% black. Despite the fact that whites comprise fewer than 9 percent of Illinois's charter school population, however, the average white charter student's school is 54% white and only 34% black (Table 8).

Latino charter school enrollment patterns are mixed. In only eight of the sixteen states analyzed, Latinos comprise a higher percentage of the total charter school enrollment than the state's public school enrollment. In some states (Texas, California, Arizona, and Pennsylvania) there are relatively high levels of Latino isolation for the average Latino charter school student. Latino students in Texas experience the highest isolation of all Latino charter students with the typical Latino charter student attending schools where two-thirds of the student body is Latino. In most states, however, Latinos are less racially isolated than either black or white charter school students.

Charter vs. public schools

Not surprisingly, given the higher proportion of black students enrolled in charter schools when compared to public schools, the average black charter school student is more isolated than his or her public school counterpart. Georgia is the only state (Table 9) in which black students are less isolated, on average, in charter schools than in public schools. This could be due to the fact that of all states, Georgia has the lowest percentage of charter school students in central cities (see Table 6). Whereas in eight states, black public school students attend schools where black students compose more than half of the student body, the typical black charter school student attends a majority black school in fourteen states.

The pattern of segregation for Latino charter school students is more varied: in six states (Florida, Massachusetts, Michigan, Minnesota, Missouri, and Pennsylvania), Latino students are more isolated in charter schools than in public schools. For example, the schools of the average Latino charter school student in Minnesota have three times as many Latino students as those of their Latino public school counterparts. However, there are also eight states in which Latino charter students are less isolated than Latino public school students. Overall, there is no clear pattern for Latino charter school student isolation.

Table 9
Minority Isolation in Public and Charter Schools, by Race/Ethnicity and by State, 2000-01(Ranked by Percent White of Charter School Students)

State	Black Isolation (Black/Black Exposure)			Latino Isolation (Latino/Latino Exposure)	
	Charter Schools	Public Schools		Charter Schools	Public Schools
Illinois	77	70		43	55
Missouri	88	61		21	10
New Jersey	83	52		34	45
Texas	72	40		66	66

Ohio	88	63		5	14
Pennsylvania	86	60		52	34
Michigan	86	74		25	22
California	51	24		55	63
Wisconsin	73	58		24	25
Florida	66	48		49	46
Minnesota	70	32		39	13
North Carolina	76	48		7	11
Massachusetts	60	34		43	39
Arizona	26	9		52	57
Georgia	55	64		13	19
Colorado	30	25		31	43

Minority Student Exposure to White Students

Within the charter school sector

Given the relatively high percentage of black students in charter schools and the levels of black isolation, we would expect to see low black exposure to whites in charter schools. It is surprising that even in states where there are more white students than black students in charter schools (e. g. , California, Wisconsin, North Carolina, Massachusetts, and Florida), the average black charter school student still attends schools with *three to four times* more black students than whites (see Table 9 and Table 10). For example, in Minnesota, where black students comprise fewer than one-quarter (23%) of all charter students (Table 6), the average black charter school student attends a school that is 70% black (Table 9) and only 17% white (Table 10). The average white student attends a school with a very different racial composition, one that is 83% white and only 7% black (Table 8).

In every state, Latino charter school students experience similar, or greater, exposure to white students when compared to the black charter school students. There are five states in which the average Latino charter school student attends school where there are at least 40% white students. While this could be due to the fact that in most states there is a much larger black share of charter school students than Latino students, we see that in states such as Texas, where the Latino and black student composition is similar (37% and 41%, respectively), the average Latino student attends a school with a greater percentage of white students (14% versus 9%, respectively).

Charter vs. public schools

When comparing charter schools to public schools for minority students, in every state except Georgia (which as mentioned earlier, is the only state in which black charter student isolation is lower than black public student isolation), black exposure to whites is higher in public schools than in charter schools (see Table 10). Black students in the public schools of these fifteen states, on average, attend schools with substantially higher proportions of white students than the average black charter school student. In Illinois, both black charter and public school students have the lowest exposure to whites than in any of the sixteen states. Whereas the state's black public school student exposure to white students is just under 20%, black charter school student exposure to white students is only 4%.

In eleven states, the average Latino student has lower exposure to white students in charter schools than public schools, and in some states they are substantially less exposed to white students than in public schools. For example, in Texas, a Latino public school student attends a school, which is, on average, 23% white; the typical Latino charter school student in this state has just over 14% white students in his or her school.

Table 10
Minority Student Exposure to White Students in Charter and Public Schools, by Race/Ethnicity and State, 2000-01 (Ranked by Percent White of Charter Students)

State	Black/White Exposure			Latino/White Exposure	
	Charter Schools	Public Schools		Charter Schools	Public Schools
Illinois	4	19		3	29
Missouri	7	35		11	69
North Carolina	21	44		40	53
Texas	9	29		14	23
Ohio	11	34		35	64

Pennsylvania	9	30		16	41
Michigan	11	21		43	59
California	16	24		25	21
Wisconsin	16	30		40	54
Florida	25	36		34	33
Minnesota	17	45		31	66
New Jersey	5	26		12	29
Massachusetts	22	40		26	40
Arizona	35	44		34	33
Georgia	36	31		49	46
Colorado	39	44		54	46

In summary, the exposure and isolation indices suggest that, due to the disproportionately high enrollment of blacks and under-enrollment of white students in charter schools when compared with public school enrollment, the average white charter student attends a school with more minority students than the average white public school student. Conversely, because of the small proportion of whites, the average black— and to a certain extent, the average Latino— student is generally more isolated in charter schools than in public schools. Although white isolation among public school students is the highest, among charter school students, black isolation is as high as white isolation. Even in states in which white enrollment is higher than black enrollment in charter schools, blacks still attend schools with three to four times the number of white students. Latino charter school student segregation from white students is lower than that of black charter students, and is not uniformly more segregated in comparison to public school students in these states. While Latino charter student exposure to whites is higher than blacks, in most states it is still lower than that of Latino public school students.

Racial Segregation

Because the exposure index only shows what the average student experiences, we now turn to other segregation measures that examine how students are distributed across schools. To gain a clearer picture of the distribution of charter school students, we examine the percentage of students of each race that attend predominantly minority schools, intensely segregated minority schools, and intensely segregated white schools.

Predominantly Minority Charter Schools

Within the charter school sector

The proportion of white, black, and Latino students attending charter and public schools where more than 50% of the student body is minority is presented in Table 11. The white share of enrollment in both charter and public schools—as well as the difference in white enrollment between the two— are in columns 1 to 3. For example, 9% of Illinois's charter school enrollment is white and 60% of its public school enrollment is white, a difference of 51 percentage points. Columns 4 to 6 show the percent of white, black, and Latino students who are enrolled in charter schools that are predominantly minority. In Illinois, 32% of white charter school students, 98% of black charter students, and almost all Latino charter school students are enrolled in 50-100% minority schools. As columns 7 to 9 show, 8% of whites, 82% of blacks, and 74% of Latinos attend 50-100% minority public schools in Illinois. Regardless of race, a higher percentage of charter school students attend predominantly minority schools when compared to public school students, which is not surprising given the much smaller percentage of white students in charter schools than in public schools in Illinois.

As discussed above, charter schools in twelve of the sixteen states enroll, in aggregate, a lower percentage of white students than public schools. In some states, these differences are stark. As column 4 shows, low percentages of white students in many states attend predominantly minority charter schools, regardless of the white share of enrollment. For example, charter school enrollment in Pennsylvania is 30% white, yet only 13% of white charter school students attend predominantly minority schools (Table 11). In fact, ten of the sixteen states have fewer than one-fifth of white charter school students attending predominantly minority schools. However, there are variations. In New Jersey, white charter school students are exposed to large proportions of students from other racial groups: 61% of white charter students in the state attend predominantly minority charter schools. This could be due to the small percentage of whites in charter schools (12%). Yet, in Illinois, a state with smaller proportion of white students in charter schools (9%), only 32% of white students attend predominantly minority charter schools, which seems to indicate that the charter school segregation of whites and blacks in Illinois is more extreme than in New Jersey.

For black students there is less variation in the percentage attending predominantly minority schools: in virtually every state there is a majority—and often an overwhelming majority—of black charter school students that attend schools with at least 50% minority students, regardless of the white proportion of the state's charter school enrollment (see column 5). In fact, in half of the sixteen states, over 90% of black students attend predominantly minority schools. This may be due to the low white charter enrollment in some of these states. However, even in some states where at least half of the charter school population is white (e. g. , Arizona, Georgia, Massachusetts, North Carolina, and Minnesota), at least three out of every five black students attend predominantly minority charter schools. In the case of Colorado, where 74% of the charter school enrollment is white, almost 60% of black students attend predominantly minority schools. Minnesota and North Carolina, states with racial guidelines in their charter legislation and where over 50% of the charter school enrollment is comprised of whites, have an overwhelming percentage of black students attend predominantly minority charter schools (91% and 83%, respectively).

As shown in column 6, Latino-white charter school segregation is less severe than black-white student segregation but is still high. For example, 60% of Latino charter students in North Carolina attend predominantly minority schools, whereas 83% of black charter school students attend such schools. By contrast, only 11% of white charter school students attend predominantly minority schools. Except in two states (Georgia and Colorado), at least half of Latino charter school students are in predominantly minority schools. In most states, however, a lower share of Latino charter students are in predominantly minority schools than are black charter school students.

Charter vs. public schools

Comparing the enrollment rates of predominantly minority charter schools (columns 4 to 6) to that of predominantly minority public schools (columns 7 to 9) illustrates that in a majority of states, regardless of race, students are more likely to attend predominantly minority charter schools than predominantly minority public schools. This is especially true for black students. A higher proportion of blacks attend predominantly minority charter schools than public schools in all except two states (Georgia and Colorado). For Latino students, this is true in all except five states (California, Florida, Arizona, Georgia, and Colorado) (see Table 11). One possible explanation could be the relatively higher enrollment of white students in charter schools in these states by comparison to other states. In most states with lower white charter school enrollment than white public school enrollment, a higher percentage of white charter school students than white public school students are enrolled in predominantly minority schools.

Table 11
Percentage of Charter and Public School Students in Predominantly Minority Schools by Race/Ethnicity and by State, 2000-01(Ranked by Percent White of Charter School Students)

State	White Share of School Enrollment			50-100% MinorityCharter School Enrollment Rate			50-100% MinorityPublic School Enrollment Rate		
	Charter (1)	Public (2)	Charter-Public Difference (3)	White (4)	Black (5)	Latino (6)	White (7)	Black (8)	Latino (9)
Illinois	9	60	-51	32	98	100	8	82	74
Missouri	9	80	-71	100	100	100	3	66	26
New Jersey	12	61	-49	61	98	98	9	75	74
Texas	20	42	-22	48	97	95	24	75	83
Ohio	25	81	-56	30	94	69	4	70	33
Pennsylvania	30	79	-49	13	92	85	3	70	64
Michigan	40	75	-35	12	91	57	3	82	39
California	42	36	6	23	88	82	34	86	87
Wisconsin	48	81	-33	17	89	71	3	72	42
Florida	50	54	-4	17	79	68	16	64	71
Minnesota	52	83	-31	12	91	72	4	59	30
North Carolina	53	61	-8	11	83	60	16	59	45
Massachusetts	54	76	-22	18	89	83	6	67	64
Arizona	56	53	3	10	62	65	17	55	72
Georgia	64	55	9	10	67	50	16	72	56
Colorado	74	68	6	4	56	35	10	57	54

Intensely Segregated Minority Schools

Within the charter school sector

Examining the distribution of students in intensely segregated minority schools, it becomes even more apparent how isolated minority students are in charter schools. (Note 44) The percentage of white, black, and Latino students that are attending charter and public schools where more than 90% of the student body is minority is shown in Table 12. Columns 1 to 3 show the white share of enrollment in both charter and public schools. The percentage of white, black, and Latino students who are enrolled in intensely segregated minority charter schools are in columns 4 to 6, and the share of students by race enrolled in intensely segregated minority public schools are in columns 7 to 9. For example, Massachusetts, a state where white students comprise 54% of total enrollment in its charter schools, has 2% of white charter students, 56% of black charter students, and 40% of Latino charter students attending intensely segregated minority charter schools. In Massachusetts' publi schools, which have a greater percentage of white students enrolled compared to charter schools, a lower percentage of all students are in intensely segregated minority schools. Less than one-half of one percent of white public school students, twenty-three percent of black students and 18% of Latino students are attend these intensely segregated minority schools.

As column 4 shows, low percentages of white charter students are in intensely segregated minority charter schools. Except in three states (Illinois, Missouri, and New Jersey), fewer than 10% of white students in charter schools attend 90-100% segregated minority schools. Even in states where the white share of charter enrollment is very low, such as Illinois (9%) and Missouri (9%), only 25% and 21% of white students, respectively, attend these intensely segregated minority charter schools (see Table 12). However, if students were evenly distributed in Illinois charter schools, for example, every school would be 9% white and thus all white charter students (as well as all minority charter school students) would be attending the intensely segregated minority schools.

In every state except Arizona, Georgia, and Colorado, at least half of black charter school students attend 90-100% minority schools (see column 5 in Table 12). A striking example is Pennsylvania, where 80% of black charter school students attend intensely segregated minority schools.

Latino charter school students experience higher segregation than that of whites and lower segregation than blacks (column 6). Five states have more than half of Latino charter school students in intensely segregated minority schools. However, except for Illinois, the attendance of Latino students at 90-100% minority schools, while still high, is less severe than that of blacks. In Minnesota, the first state to enact a charter law (which includes racial/ethnic balance guidelines), and a state with very high white charter school enrollment, demonstrates high levels of charter segregation for minority students with roughly two out of every three black and two out of every five Latino charter school students attending intensely segregated schools.

Part of this segregation may be due to the higher percentage of minority students enrolled in charter schools, which results in more predominantly minority schools. But the racial disparities among these schools suggest that there is another factor aside from the racial composition in the state's charter schools that is driving these numbers. For example, as we have seen earlier, even if they are a small proportion of students in charter schools, whites are not as likely as black and Latino students to attend heavily minority schools. This indicates that the over-enrollment of minority students in charter schools is more likely to result in highly segregated schools for minorities than for whites. These trends of disproportionately high enrollment of minority students in intensely segregated schools could also be due to the fact that some of the charter schools are located in segregated central city neighborhoods. It is worth remembering, however, that charter schools as schools of choice are not limited to neighborhoods or even public school districts, but can draw students from a larger geographical area.

Charter vs. public schools

Charter school students across all racial groups in most of the sixteen states are more likely to attend intensely segregated minority schools than are public school students (see columns 4 to 9). In both sectors, however, attendance at such schools differs substantially by race. In every state, a higher percentage of black students in charter schools than in public schools are enrolled in intensely segregated schools (see Table 12). In California, Arizona, and Texas, the three states with the largest charter school enrollment, black charter school students are attending intensely segregated charter schools at rates almost two times higher than black public school students. In Illinois, Missouri, New Jersey, Pennsylvania, Minnesota, and North Carolina, the share of black students attending intensely segregated charter schools is more than thirty percentage points greater than those in intensely segregated minority public schools. Of these states, North Carolina, Missouri, New Jersey, and Minnesota have racial guidelines in their charter legislation.

In all states, regardless of the type of school (i. e. , charter or public), fewer than 25% of white students attend 90-100% minority schools. It is worth noting that a higher percentage of white charter school students than white public school students are in intensely segregated minority schools in twelve of sixteen states. In three states, fewer than one percent of white charter school students are in 90-100% minority schools.

Table 12
Percentage of Charter and Public School Students in Intensely Segregated Minority Schools, by Race/Ethnicity and by State, 2000-01 (Ranked by Percent White of Charter School Students)

State	White Share of School Enrollment			90-100% MinoritySchool Enrollment Rate			90-100% MinorityPublic School Enrollment Rate		
	Charter (1)	Public (2)	Charter-Public Difference (3)	White (4)	Black(5)	Latino (6)	White (7)	Black (8)	Latino (9)

Illinois	9	60	-51	25	94	98	1	60	40
Missouri	9	80	-71	21	77	63	0	35	3
New Jersey	12	61	-49	13	87	76	1	49	41
Texas	20	42	-22	8	74	63	2	36	47
Ohio	25	81	-56	6	74	29	0	34	3
Pennsylvania	30	79	-49	2	80	69	0	47	27
Michigan	40	75	-35	1	77	10	0	61	10
California	42	36	6	1	68	43	2	36	44
Wisconsin	48	81	-33	1	63	3	0	42	17
Florida	50	54	-4	1	53	27	1	30	30
Minnesota	52	83	-31	1	65	39	0	15	4
North Carolina	53	61	-8	1	60	41	0	10	4
Massachusetts	54	76	-22	2	56	40	0	23	18
Arizona	56	53	3	1	28	26	1	12	26
Georgia	64	55	9	1	41	14	1	35	13
Colorado	74	68	6	0	35	16	0	19	15

Intensely Segregated White Schools

Within the charter school sector

Table 13 displays the percentage of students by race that attends intensely segregated white charter schools. (Note 45) Despite relatively low white charter school enrollment rates, there are only 2 states (Illinois and Missouri) without any students attending intensely segregated white charter schools (see columns 4 to 6), and in some states, white isolation is particularly stark. For example, despite the fact that about 60% of Michigan's charter school students are minority, 40% of white students attend intensely segregated white charter schools (see Table 13). In fact, in 10 states at least 15% of white charter school students attend intensely segregated white schools; in six states, over one-quarter of all white charter school students are in intensely segregated white schools.

In every state except Massachusetts, blacks are the least likely of all students to enroll in intensely segregated white charter schools. (Note 46) In no state are there greater than 4% of black students in intensely segregated white charter schools and, further, regardless of the white share of total enrollment, fewer than 10% of black students—public or charter— are enrolled in intensely segregated white schools in all states (column 5).

When compared to black students, higher percentages of Latino students are in intensely segregated white charter schools but still fewer than 10% of Latino charter school students in every state attend such schools (column 6).

Charter vs. public schools

In most states, a lower percentage of white charter school students attend intensely segregated white schools than white public school students, which would be expected given the lower percentage of white students in charter schools. There are five states in which a higher proportion of white charter school students by comparison to public school students attend90-100% white schools (i. e. California, Florida, North Carolina, Arizona, and Colorado). Interestingly, in North Carolina, a state with racial/ethnic balance guidelines in their charter legislation, and where white share of the charter school enrollment (53%) is smaller than the public school enrollment (61%), there is a *higher* percentage of white charter school students in intensely segregated white schools than white public school students, indicating that, on average, white students in charter schools are more isolated than in the public schools in North Carolina.

For minority students, in states where there is a higher percentage of black students in 90-100% white charter schools than in the public schools (i. e. , Arizona, California, Georgia, and Colorado), there is still only a very small presence of black students in intensely segregated white schools. For Latino students, there are six states in which more than 10% of Latino public school students are enrolled in 90-100% white schools, but Latino charter school students are generally less likely to attend intensely segregated white schools than Latino public school students. However, in four states (i. e. California, Arizona, Florida, and Colorado), a higher share of Latino charter school students are in intensely segregated white schools than are Latino public school students.

In general, minority students in charter schools are less likely to be in heavily white schools than minority students in public schools. This could be due to the larger enrollment share of minority students in charter schools. In most states, white charter school students are less likely to be in intensely segregated white schools than public school students, but on average, they enroll in intensely segregated white charter schools at rates much higher than black and Latino charter school students.

Table 13
Percentage of Charter and Public School Students in Intensely Segregated White Schools, by Race/Ethnicity and by State, 2000-01 (Ranked by Percent White of Charter School Students)

State	White Share of School Enrollment			90-100% White Charter School Enrollment Rate			90-100% White Public School Enrollment Rate		
	Charter (1)	Public (2)	Charter-Public Difference (3)	White (4)	Black (5)	Latino (6)	White (7)	Black (8)	Latino (9)
Illinois	9	60	-51	0	0	0	48	2	3
Missouri	9	80	-71	0	0	0	69	6	32
New Jersey	12	61	-49	8	0	0	35	2	3
Texas	20	42	-22	6	0	0	10	0	1
Ohio	25	81	-56	2	0	0	76	6	28
Pennsylvania	30	79	-49	26	0	1	74	7	11
Michigan	40	75	-35	40	1	7	71	4	25
California	42	36	6	8	0	0	4	0	0
Wisconsin	48	81	-33	39	1	4	67	4	18
Florida	50	54	-4	17	1	2	14	1	1
Minnesota	52	83	-31	60	2	9	65	8	22
North Carolina	53	61	-8	23	1	4	20	1	5
Massachusetts	54	76	-22	54	3	2	61	8	6
Arizona	56	53	3	16	1	2	8	1	1
Georgia	64	55	9	21	1	3	22	1	6
Colorado	74	68	6	37	4	7	23	2	3

In earlier sections of this article, analysis using the exposure index showed that the average white charter school student was less segregated from minorities than the average white public school student. Conversely, black charter school students are more isolated than their public school counterparts, and the record was mixed for Latino charter school students. In this section, we examine the distribution of races within predominantly minority, intensely segregated minority, and intensely segregated white schools.

Students of all races are more likely to enroll in predominantly and intensely segregated minority charter schools than their public school counterparts. However, the percentages of white students in these minority charter schools were still much lower than those for black and Latino students. We speculate that the over-enrollment of minority and under-enrollment of white students in charter schools might result in more students attending predominantly minority and intensely segregated minority schools. This over-enrollment of minority students should make it possible to expose whites (as well as black and Latino students) to greater percentages of minority students, and we have seen that, in fact, white charter school students in many states are less isolated than their public school counterparts. However, given the high white isolation of public school students, white charter school students are still heavily enrolled in intensely segregated white schools. Even in states with a predominantly minority population in their charter school population, few white charter school students attend heavily minority charter schools. Thus the possibility of substantial interracial exposure of white students to minority students is largely unrealized. In most of these sixteen states, black and Latino charter school students are attending segregated minority schools at an even higher rate than those in the increasingly resegregating public schools.

Conclusion

The driving idea behind the charter school movement has been allowing schools greater autonomy in exchange for greater accountability. After a decade of rapid expansion and huge increase in public support for charter schools, often on the basis of arguments that they improve equity in school systems, it is time to hold these schools accountable for their accomplishments.

Our study shows that charter schools face high levels of segregation. Certainly there is tremendous variation among schools: some are highly diverse while others have high levels of isolation, particularly for black students. Although these schools have the potential to transcend high residential segregation created by neighborhood assignment and school district boundary lines, in many cases they are even more segregated than regular public schools. This might be due to the fact that many charter schools are located in segregated neighborhoods; establishing charter schools on boundaries between white, black and Latino neighborhoods could increase the likelihood of drawing a diverse student body.

Our state data suggest that black students are enrolled in charter schools—as well as intensely segregated minority charter schools— at a rate nearly twice their share of the public school population. Despite higher minority enrollments in charter schools, however, we still see in a number of states that whites are racially isolated. We find that regardless of white share of the entire charter school enrollment, black students in charter schools experience high levels of racial isolation and are exposed to very low percentages of white students. There is little evidence from this analysis that the existence of charter schools helps to foster more integrative environments, especially for minority students. At a time when the public schools are more segregated for minority students than thirty years ago, any reform that is publicly funded and intensifying the increasing public school segregation deserves very careful evaluation.

We continue to learn about the benefits of racial and ethnic diversity in schools for students of all races and at the same time, according to public opinion polls, public support for racial diversity is increasing. (Note 47) Further in a recent case concerning affirmative action in higher education, *Grutter* v. *Bollinger*, the Supreme Court recognized the importance of diversity as a compelling state interest. This article shows that instead of creating schools of diversity, many charter schools are places of racial isolation, particularly for minority students. Based on lessons learned in other school choice programs, such as magnet schools, the following conditions may help to address issues of racial isolation by creating a system that allows students to choose to attend charter schools on an equitable basis:

1. Full information: The theory of choice as an equitable system has always depended on full information to all families. Information about charter schools and application procedures are often linked to social networks. Information must be made available to all potential students and parents, and in a language that all can understand. This might be aided by centralizing means of charter information dispersal in state departments of education and/or charter offices regardless of which agencies and organizations are allowed to grant charters.
2. The provision of free transportation for all students, even across school district boundaries, is essential to ensuring that all interested students can choose to attend charter schools. Students of poorer families will see their opportunities to choose constrained where charter schools are not required to provide transportation.
3. Providing for and welcoming all groups, including students from all racial/ethnic groups, English Language Learners, and special education students. In many ways, both implicitly and explicitly, charter schools can make their environment unwelcoming for a diverse array of students. Simply put, *any* publicly funded school should be a place where all students could be effectively educated.
4. No screening of children for charter schools, both academic and otherwise. Although most states require that charter schools enroll students on a first-come, first-serve basis, legislation in some states allows schools to employ both academic and non-academic criteria in student enrollment. Admissions procedures that might unfairly prohibit any child from enrolling (such as pre-admissions interviews or a requirement of parental involvement in the school) should be eliminated. Some states, such as Michigan, have tried to address this by specifying that admissions processes be made public.

No Child Left Behind provides an opportunity for all students in low-performing schools to attend better schools, including moving to charter schools. We believe that this transfer opportunity should include a majority- to-minority transfer to all charter and magnet schools where room is available, and that the transfer will increase racial integration in the sending and receiving schools. As such, transportation should be provided for students across a metropolitan area.

To ensure that choice policies and charter schools promote racial equity and integrated schools, a number of political scientists and policymakers have underscored the need for government regulation of education markets (Cobb & Glass, 1999; Moe, 2002; Taebel et al, 1997). For example, Hill and Guin (2002) assert that "choice programs must be carefully designed to prevent segregation, and any program that produces levels of segregation as great as those now prevailing in the public education system should be scrapped or redesigned" (p. 49). Our findings suggest that many state charter laws need to be redesigned to include stronger enforcement mechanisms to ensure racial integration. State education agencies should develop policies to ensure that the four conditions above exist wherever charter schools are authorized. They should provide support and encouragement for schools to create a diverse student body and to recruit students of all races. Indeed, charter laws should incorporate lessons learned from regulated choice plans, such as controlled open enrollment and magnet schools, that have produced stable, integrated schools in many districts including Minneapolis, Minnesota and Cambridge, Massachusetts (Willie, 2000). On the other hand, permissive charter school laws and unregulated choice policies have increased racial isolation for black students and facilitated white flight from integrated schools in Arizona (Cobb & Glass, 1999). Given the increasing ethnic separation in Arizona charter schools, Cobb and Glass (1999) argue that charter schools "should be required to actively pursue ethnic representation" (p. 31).

If charter schools are to be an educational reform that provides an alternative means to broaden access to high quality education, issues of racial/ethnic segregation and practices that create the disturbing patterns of racial isolation in charter schools in many of our states, as detailed in this article, must be closely examined. In addition to monitoring student achievement and financial management, charter granters must hold charter schools to racial/ethnic balance guidelines in those states and districts with such legislation or court orders. Ultimately, the extent of public oversight over school choice will determine, to a large extent, whether charter schools support or undermine racial integration in public education.

Notes

1. We would like to thank Gary Orfield for his leadership at the Civil Rights Project and assistance on this project. Our thanks also go to Catherine Horn, Al Kauffman, Jimmy Kim, Michal Kurlaender, Patricia Marin, and John Yun of The Civil Rights Project for their invaluable comments and suggestions as well as Tiana Davis and the staff at the Civil Rights Project for their assistance. Special thanks to external reviewers for their feedback, from which this article also benefited: Carol Ascher, New York University; Casey Cobb, University of Connecticut; Gene Glass, Arizona State University; Jerry Horn, Western Michigan University; Gary Miron, Western Michigan University; Janelle Scott, New York University; and Amy Stuart Wells, Teachers College, Columbia University as well as anonymous reviewers from this journal.
2. Teachers in charter schools, however, also reported serious frustrations and difficulties in environments lacking in security, clear authority, career development, and other stresses. (Susan Moore Johnson and Jonathan Landman, "'Sometimes Bureaucracy Has Its Charms': The Working Conditions of Teachers in Deregulated Schools," Teachers College Record, vol. 102, no. 1 (February 2000), pp. 85-124).
3. Lee McGraw Hoffman, Overview of Public Elementary and Secondary Schools and Districts: School Year 2001-02, National Center for Education Statistics, May 2003, table 9, p. 21.
4. Karla Scoon Reid, "Minority Parents Quietly Embrace School Choice," Education Week, December 5, 2001.
5. Erica Frankenberg, Chungmei Lee & Gary Orfield, A Multiracial Society with Segregated Schools: Are We Losing the Dream?, Cambridge, MA: The Civil Rights Project at Harvard University, 2003.
6. Ellen Sorokin, "Poll finds most blacks favor charter, private schools," The Washington Times, July 19, 2002; Sharon Terlep, "Charter school study finds support is solid, MSU report may play role in effort to lift current cap," Lansing State Journal, December 15, 2002.
7. Tom Loveless, How Well are American Students Learning? Part III, "Charter Schools," The 2002 Brown Center Report on American Education, Washington: Brookings Institution, 2002, pp. 30-36; David Arsen, David N. Plank, and Gary Sykes, "A Work in Progress," Education Next, vol. I, No., 4, winter 2001, pp. 14-19; a large Texas study found no difference (Eric A. Hanushek, John F. Kain, and Steven G. Rivkin, "The Impact of Charter Schools on Academic Achievment," unpublished report, December 2002).
8. Amy S. Wells and Robert L. Crain, Stepping Over The Color Line: African-American Students in White Suburban Schools. New Haven: Yale University Press, 1997; Bruce Fuller and Richard Elmore (eds.); Who Chooses, Who Looses? New York: Teacher College Press, 1996.
9. Wendy S. Grigg, Marcy C. Daane, Ying Jim and Jay R. Campbell, "The Nation's Report Card: Reading 2002," National Center for Education Statistics, June 2003.
10. For data from Fordham Foundation, an active charter school supporter, suggesting weak oversight in many states, see, "Grading the Chartering Organizations," Education Week, June 11, 2003.
11. Bruce Fuller and Richard Elmore (eds.), Who Chooses, Who Looses? New York: Teacher College Press, 1996.
12. Statutes concerning charter schools are found using Westlaw and Lexis-Nexis. The specific citations are available upon request from the authors.
13. Unless otherwise indicated, all the authors' tabulations are from the 2000-01 NCES Common Core of Data.
14. See Table 6 infra.
15. No Child Left Behind Act of 2001 § 1116(b)(1)(E).
16. The FY02 federal budget allocated $200 million in competitive grants for "expanding the number of high-quality charter schools available to students across the Nation" (No Child Left Behind Act of 2001 § 5201(3)).
17. See Table 9 in Frankenberg, Lee, and Orfield (2003).
18. For example, of the 1855 schools, only 291 schools reported free and reduced lunch data. Of these, 63% of the schools had student bodies with than 10% black and Latino students. While it is interesting to note that segregated white charter school are more likely to offer the free and reduced lunch program than other charter schools, these data are not reliable enough to draw any conclusions as to the correlation of racial minority and poverty concentration.
19. Note: The CCD racial categories, as derived from information submitted by each state, are: non-Hispanic White, non-Hispanic Black, Hispanic, Asian, and Native American. Thus, our analysis is limited to these categories and cannot include, for example, biracial students.
20. NCES defines a charter school as, "a school that provides free elementary and/or secondary education to eligible students under a specific charter granted by the state legislature or other appropriate authority." http://nces.ed.gov/ccd/data/txt/psu00lay.txt
21. The term isolation is used to denote the exposure of one race to itself, for example, white to white. This is another measure of segregation, which shows how concentrated students are with other students of their own race. We use the terms white-white exposure and white isolation interchangeably throughout the report to refer to the exposure of white students to other white students in their school.
22. There are certain shortcomings to comparing individual charter schools to district averages since these averages are, in general, more diverse than individual non-charter public schools (Wells, Holme, Lopez, and Cooper, 2000).
23. There is some evidence that supports the idea that charter schools are attracting students from a broader geographic area than other public schools. In Pennsylvania, Miron, Nelson, and Risley (2002)

found that charter school students traveled an average of 5.6 miles from their home to charter school whereas other public school students traveled 2.4 miles. In theory, local districts are responsible for transportation arrangements, yet Miron and colleagues note that that some districts are still working out these details. Miron and Horn (2002) found similar patterns of longer distances to charter schools than traditional public schools in Connecticut as well.

24. 20 U.S.C. 8062 (1994).
25. 20 U.S.C. 8061 (1994). For a more detailed treatment on the civil right provisions and charter schools, see Wohlstetter et al., (1995).
26. Wright v. Council of Emporia, 407 U.S. 451, 460-462 (1972) (a new school district could not be created if its effect would be to impede progress of dismantling an existing dual system). Also, for more recent cases that specifically pertain to charter schools, see Berry v. School District of the City of Benton Harbor, 56 F.Supp.2d 866, 872 (W.D. Mich. 1999) (when considering charter school application to operate within a dual school system, court will consider interference with remedial order and effect on court's ongoing ability to eliminate vestiges of discrimination); Beaufort County Bd. of Educ. v. Lighthouse Charter School et. al., 516 S.E.2d 655, 659 (S.C. 1999) (upholding a school board finding that a prospective charter school failed to adhere to same reporting requirements under OCR Title VI desegregation plan as other public schools in the district); Davis v. East Baton Rouge Parish School Board, et al, C.A. No. 56-1662 (M.D. La. 1999) (stating that charter schools in district remain subject to court's orders relating to desegregation of district).
27. Essentially, the 2000 U.S. Department of Education guidelines only tell prospective charter school founders to determine whether their proposed school is in a district with a school desegregation plan, and, if so, to consult with Department of Education officials. (See Parker, 2001). Note: Some states are now starting to address this in revising their charter laws.
28. Statutes governing charter schools are found using Westlaw and Lexis-Nexis. The specific citations of statutes are available upon request from the authors.
29. The states comprising our definition of the South, as traditionally used in documenting school segregation trends, are the former slave states that practiced legally mandated segregation: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia. Our definition of other regions is as follows: Border: Delaware, Kentucky, Maryland, Missouri, Oklahoma, and West Virginia; Northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont; Midwest: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin; West: Arizona, California, Colorado, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming. Note: Alaska and Hawaii are excluded because of their unique ethnic compositions and isolation from the regions studied here.
30. This section analyzes data from the 34 states with operational charter schools. Thus, public school trends in these 34 states may be slightly different than national trends based on all 50 states.
31. Throughout this report, in all data presented in tables comparing public and charter schools, we have removed charter schools from the public school data. Therefore, we can compare charter schools with non-charter public schools.
32. For the remainder of the report, we use the term "predominantly minority" to designate schools where at least 50% of the student body is minority. Likewise, we use the term "intensely segregated minority" to designate schools where at least 90% of the student body is minority.
33. Of course, at least in intensely segregated minority schools, by definition there will be a small percentage of white students.
34. To compare to charter student exposure in Table 6 of this report, see Table 4, page 27 in Frankenberg, Lee, & Orfield (2003).
35. Due to the small numbers of Asian and Native American students in charter schools in most states (although there are exceptions such as Minnesota and Arizona), the state-level analysis of racial/ethnic segregation will not include these students.
36. The Common Core of Data has eight categories for locale: large city, mid-size city, urban fringe of large city, urban fringe of mid-size city, large town, small town, rural outside metropolitan statistical area (MSA), and rural inside MSA. We defined the three categories of urban, suburban, and rural based on NAEP's definitions. As defined by NAEP, central cities include all central cities in Standard Metropolitan Statistical Areas (SMSAs) as determined by the Office of Management and Budget. Urban Fringe/Large Town denotes large towns that are located within SMSA's that are urban but not defined as central city. Rural/Small Town areas include all areas that are classified as rural by the Census. For the purposes of this report, we will use central cities, suburban for urban fringe or large town areas, and rural for small town and rural areas.
37. For data on the racial composition of the largest public school districts, see Frankenberg, Lee, and Orfield (2003), pp. 53-57.
38. The states where charter schools over-enroll white students are either in the South or West (see Figure 2). One reason suggested for this trend is that in states with large and/or diverse public school systems, charter schools might provide a means for white students to avoid racially diverse schools (Wells, et al. 2000). The South and the West are also the two regions of the country with the highest percentages of minority public school students, which are almost 50% (Frankenberg, Lee, and Orfield, 2003).
39. It should be noted, however, that Missouri only authorizes charter schools in St. Louis and Kansas City. As these urban areas are heavily minority, it is not surprising that Missouri charter schools enroll such a high percentage of African-American students.
40. There are exceptions to this trend for white public school students in four states (Colorado, Arizona, Florida, and California) in which white isolation is actually lower than the white percentage of the state's total enrollment.
41. E.g., the difference in the white enrollment share is less than ten percentage points.

42. The average white public school student attends a school that is 79.7% white. (Frankenberg, Lee & Orfield, 2003).
43. Racial isolation also has a high correlation with student poverty; of all public schools nationwide, 86% of schools in 2000-01 that had 90-100% minority students were schools in which at least half the student body was poor or near poor (Frankenberg, Lee & Orfield, 2003).
44. Intensely segregated white schools tend to be schools with a lower percentage of poor or near poor students; nationally, less than 15% of schools that are 90-100% white are likely to be schools of concentrated poverty (Frankenberg, Lee & Orfield, 2003).
45. Latino students in Massachusetts are enrolled in intensely segregated white schools at a lower percentage than blacks (2% for Latino students to 3% for black students).
46. See discussion supra.

References

20 U. S. C. 8061 (1994)

20 U. S. C. 8062 (1994)

American Institutes for Research. (February 1993). *Magnet Schools and Issues of Desegregation, Quality and Choice*, Phase I: The National Survey and In-Depth Survey of Selected Districts.

Apple, M. (2001). *Educating the "right" way: Markets, standards, God and inequality*. New York: Routledge.

Arsen, D., Plank, D. N., & Sykes, G. (1999). *School choice policies in Michigan: The rules matter*. East Lansing: Michigan State University Press.

Ascher, C., Jacobowitz, R., & McBride, Y. (1999). *Standards-based reform and the charter school movement in 1998-99: An analysis of four states*. New York University, Institute for Education and Social Policy.

Beaufort County Board Of Education v. Lighthouse Charter School et al., 516 S. E. 2d 655, 659 (S. C. 1999).

Berry v. School District of the City of Benton Harbor, 56 F. Supp. 2d 866, 872 (W. D. MI 1999).

Blank, R. K., Levine, R. E., & Steel, L. (1996). After 15 years: Magnet schools in urban education. In B. Fuller & R. F. Elmore with G. Orfield (Eds.) *Who chooses? Who loses?* (pp. 154-172). New York: Teachers College Press.

Braddock, II, J. H. (1980). The perpetuation of segregation across levels of education: A behavioral assessment of the contact-hypothesis. *Sociology of Education*, *53*(3): 178-186.

Clayton Foundation. (January 1999). *1998 Colorado Charter Schools Evaluation Study*. Denver, CO: Colorado Department of Education

Cobb, C. D. & Glass, G. V (1999). Ethnic segregation in Arizona charter schools. *Education Policy Analysis Archives*, *7*(1) Retrieved March, 2003 from http://epaa.asu.edu/epaa/v7n1/

Coons, J and Sugarman, S. (1978). *Education by choice: the case for family control*. Berkeley: University of California Press.

Davis v. East Baton Rouge Parish School Board, et al., C. A. No. 56-1662 (M. D. La. 1999).

Eaton, S. E. (2001). *The other Boston busing story*. New Haven: Yale University Press.

Elam, S. M., Rose, L. C., & Gallup, A. M. (1993). The 25th Annual Phi Delta Kappan/Gallup Poll of the Public's Attitude toward the Public Schools. *Phi Delta Kappan*.

Farkas, S., & Johnson, J. (with Immerwahr, S. & McHugh, J.). (1998). *Time to Move On: African Americans and White Parents Set an Agenda for Public*. New York: Public Agenda.

Frankenberg, E., Lee, C., and Orfield, G. (2003). *A multiracial society with segregated schools: Are we losing the dream?* Cambridge, MA: The Civil Rights Project at Harvard University.

Fuller, B., Gawlik, M., Gonzales, E. K., Park, S., (with Gibbings, G.). (2003). *Charter schools and inequality: National disparities in funding, teacher quality, and student support*. Policy analysis for California education, Working paper series 03-2. April, 2003, from http://pace.berkeley.edu/Chartersummary.pdf.

Fuller, B., and Elmore, R. (eds.). (1996). *Who Chooses, Who Looses?* New York:Teacher College Press.

Gallup. (1999, August). Gallup Poll Topics: Education Poll. http://www.gallup.com/

Gill, B. P., Timpane, P. M., Ross, K. E., & Brewer, D. J. (2001). *Rhetoric versus reality: what we know and what we need to know about vouchers and charter schools*. Santa Monica, CA: Rand Corporation.

Green v. County School Board of New Kent County, 391 U. S. 430 (1968).

Finn, Jr., C. E., Manno, B. V., and Vanourek, G. (2000). *Charter schools in action: Renewing public education.* Princeton, NJ: Princeton University Press.

Hill, P. T., & Guin, K. (2002). Baselines for assessment of choice programs. In P. T. Hill (Ed.), *Choice with equity* (pp. 15-49). Stanford, CA: Hoover Institution Press, p. 49.

Hochschild, J. & Scovronick, D. (2003). *The American Dream and the Public Schools.* New York: Oxford University Press.

Hoffman, L. M. (May 2003). *Overview of Public Elementary and Secondary Schools and Districts: School Year 2001-02.* Washington, D. C. : National Center for Education Statistics.

Horn, J. & Miron, G. (2000). *An evaluation of Michigan's charter school initiative: Performance, accountability, and impact.* Kalamazoo, MI: Western Michigan University, Evaluation Center.

Howe, K. R., and Welner, K. G. (2002). School choice and the pressure to perform: Deja vu for children with disabilities? *Remedial and Special Education, 23*(4): 212-21.

Kim, J. & Sunderman, G. (2003). *Findings from the first phase of school choice implementation in three districts: Buffalo, New York, Richmond, Virginia, DeKalb County, Georgia.* Paper presented at American Educational Research Association, Chicago, IL.

Kurlaender, M. & Yun, J. T. (2001). Is diversity a compelling educational interest? Evidence from metropolitan Louisville. In *Diversity Challenged*, Orfield, G., (with Kurlaender, M.) (Eds.). Cambridge, MA: The Civil Rights Project, chap 5.

Massey, D. S., & Denton, N. A. (1988). The dimensions of racial segregation. *Social Forces, 67*, 281-315.

Miron, G. & Horn, J. (2002). *Evaluation of Connecticut's Charter Schools and the Charter School Initiative.* Kalamazoo, MI: Western Michigan University, Evaluation Center.

Miron, G., & Nelson, C. (2002). *What's public about charter schools: Lessons learned aboutschool choice and accountability.* Thousand Oaks, CA: Corwin Press.

Miron, G., Nelson, C., & Risley, J. (2002). *Strengthening Pennsylvania's charter school reform: Findings from the statewide evaluation and discussion of relevant policy issues.* Kalamazoo, MI: The Evaluation Center, Western Michigan University. Retrieved March, 2003, from http://www.wmich.edu/evalctr/charter/pa_5year/

Moe, T. M. (2002). The structure of school choice. In P. T. Hill (Ed.), *Choice with equity* (pp. 179-212). Stanford, CA: Hoover Institution Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Orfield, G., Bachmeier, M., James, D., & Eitle, T. (1997). *Deepening segregation in American public schools.* Cambridge, MA: Harvard Project on School Desegregation.

Parker, W. (Feb. 2001). The Color of Choice: Race and Charter Schools. *75 Tul. L. Rev.* 563.

Public Sector Consultants & Maximus, Inc. (1999). *Michigan's charter school initiative: From theory to practice.* Lansing, MI.

Reardon, S. & Yun, J. T. (2002). *Private school racial enrollments and segregation.* Cambridge, MA: The Civil Rights Project at Harvard University.

RPP International. (2000). *The state of charter schools, national study of charter schools, fourth-year report.* Washington, D. C.: U. S. Department of Education.

SRI International. (1997). *Evaluation of Charter School Effectiveness.* Prepared for the State of California Office of Legislative Analyst. Menlo Park, CA: SRI International.

Taebel, D., Barrett, E. J., Thurlow-Brenner, C., Kemerer, F., Ausbrooks, C., Clark, C., Thomas, K., Briggs, K. L., Parker, A., Wiher, G., Matland, R., Tedin, K., Cookson, C., & Nielsen, L. (1997). *Texas open enrollment charter schools: Year one evaluation.* Texas State Board of Education.

Wells A. S. & Crain, R. L. (1994). Perpetuation theory and the long-term effects of school desegregation. *Review of Educational Research, 64*, 531-555.

Wells, A. S., Holme, J. J., Lopez, A., & Cooper, C. W. (2000). Charter schools and racial and social class segregation: Yet another sorting machine? In R. Kahlenberg (Ed.), *A notion at risk: Preserving education as an engine for social mobility* (pp. 169-222). New York: Century Foundation Press.

Wells, A. S. (Ed.). (2002). *Where charter policy fails: The problems of accountability and equity.* New York: Teachers College Press.

Willie, C. V. (2000). Controlled choice:A new and effective desegregation method. In C. C. Yeakey (Ed.), Edmund W. Gordon:*Producing* *knowledge, pursuing understanding. Advances in diverse communities:Research, policy, and praxis*, Vol. 1. Stamford, CT: JAI Press, Inc.

Wohlstetter, P. et al. (1995). Charter schools in the United States: The question of autonomy. 9 *Education Policy* 331.

Wood, J. (1999). *An Early Examination of the Massachusetts Charter School Initiative*. Amherst, MA: Donahue Institute, University of Massachusetts.

Wright v. Council of Emporia, 407 U. S. 451 (1972).

Young, B. A. and Smith, T. M. (1997). *The Condition of Education* 1997. Washington, D. C. : National Center for Education Statistics 97-991.

About the Authors

Erica Frankenberg is a Research Assistant at The Civil Rights Project and doctoral student at the Harvard Graduate School of Education. She received her Masters in Education with a concentration in Administration, Planning, and Social Policy from the Harvard Graduate School of Education. Ms. Frankenberg has also worked with a non-profit educational foundation in Alabama. She earned a B.A. in Educational Policy from Dartmouth College where she received high honors for her senior thesis regarding the end of court-mandated desegregation in Mobile, Alabama. Ms. Frankenberg's research interest in school desegregation stems from her experience as a student in desegregated public schools. She presented, "The Impact of School Segregation on Residential Housing Patterns: Mobile, AL and Charlotte, NC," at the Resegregation of Southern Schools conference in August 2002. She is also co-author of "Race in American Public Schools: Rapidly Resegregating School Districts," (with C. Lee) published in 2002 and "A Multriacial Society with Segregated Schools: Are We Losting the Dream?" (with C. Lee and G. Orfield) in 2003 by The Civil Rights Project.

Chungmei Lee is a Research Associate at The Civil Rights Project. She received her Masters in Administration, planning, and Social Policy from the Harvard Graduate School of Education. Prior to joining the Project, she worked with Harvard's Programs for Professional Education and helped train education leaders around the world in Education Management Information System (EMIS). At PPE, she also worked on issues relating to the professional development of teachers. As an independent consultant for the National Bureau of Economic Research (NBER), she examined issues such as the financing of higher education and its impact on middle-income and low-income students' access to higher education. Before coming to Harvard, Ms. Lee was a high school history teacher in a Nicaraguan international school. She holds a B.A. in history from Dartmouth College. She was born in Taiwan and lived in Honduras for many years. She is also co-author of "Race in American Public Schools: Rapidly Resegregating School Districts," (with E. Frankenberg) published in 2002 and "A Multriacial Society with Segregated Schools: Are We Losting the Dream?" (with E. Frankenberg and G. Orfield) in 2003 by The Civil Rights Project.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State Univeristy

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

PDF file of this article available:

This article has been retrieved 3145 times since September 17, 2003

Volume 11 Number 33 **September 17, 2003** **ISSN 1068-2341**

Wanted: A National Teacher Supply Policy for Education: The Right Way to Meet The "Highly Qualified Teacher" Challenge

Linda Darling-Hammond
Stanford University

Gary Sykes
Michigan State University

Citation: Darling-Hammond, L., and Sykes, G.. (2003, September 17). Wanted: A national teacher supply policy for education: The right way to meet the "Highly Qualified Teacher" challenge. *Education Policy Analysis Archives, 11*(33). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n33/.

Abstract

Teacher quality is now the focus of unprecedented policy analysis. To achieve its goals, the No Child Left Behind Act (NCLB) requires a "highly qualified teacher" in all classrooms. The concern with teacher quality has been driven by a growing recognition, fueled by accumulating research evidence, of how critical teachers are to student learning. To acquire and retain high-quality teachers in our Nation's classrooms will require substantial policy change at many levels. There exists longstanding precedent and strong justification for Washington to create a major education manpower program. Qualified teachers are a critical national resource that requires federal investment and cross-state coordination as well as other state and local action. NCLB provides a standard for equitable access to teacher quality that is both reasonable and feasible. Achieving this goal will require a new vision of the teacher labor market and the framing of a national teacher supply policy. States and local districts have vital roles to play in ensuring a supply of highly qualified teachers; however, they must be supported by appropriate national programs. These programs should be modeled on U.S. medical manpower efforts, which have long supplied doctors to high- need communities and eased shortages in specific health fields. We argue that teacher supply policy should attract well-prepared teachers to districts that sorely need them while relieving shortages in fields like special education, math and the physical sciences. We study the mal-distribution of teachers and examine its causes. We describe examples of both states and local school districts that have fashioned successful strategies for strengthening their teaching forces. Unfortunately, highly successful state and local program to meet the demand for qualified teachers are the exception rather than the rule. They stand out amid widespread use of under-prepared teachers and untrained aides, mainly for disadvantaged children in schools that suffer from poor working conditions, inadequate pay and high teacher turnover. The federal government has a critical role to play in enhancing the supply of qualified teachers targeted to high-need fields and locations, improving retention of qualified teachers, especially in hard-to-staff schools, and in creating a national

labor market by removing interstate barriers to mobility. (Note 0).

Recent policy developments have drawn unprecedented attention to issues of teacher quality. To achieve its goals for improved school outcomes, the No Child Left Behind Act (NCLB) requires a "highly qualified teacher" in all classrooms, as well as better-prepared paraprofessionals and public reporting of staff qualifications. The concern with teacher quality has been driven by a growing recognition, fueled by accumulating research evidence, of how critical teachers are to student learning. In this, policymakers have been catching up with parents, who have long believed that teachers matter most. (Note 1)

To turn the NCLB mandate into a reality, however, the nation will have to overcome serious labor market obstacles. For one, inequalities in school funding—along with widely differing student needs and education costs—produce large differentials in staff salaries and working conditions that affect the supply of teachers to different schools. For another, teacher labor markets, although starting to change, have been resolutely local. In many states, most teachers still teach in schools near where they grew up or went to college (Boyd et al, 2003). These factors, together with other labor market conditions, have meant that some schools traditionally have been "hard to staff." The hardest-hit schools chiefly serve poor, minority and low-achieving children—the same children whose learning must increase significantly if the central NCLB goal of closing the achievement gap between advantaged and disadvantaged pupils is to be accomplished. To get and keep high-quality teachers in these children's classrooms will require substantial policy change at all levels.

While more extensive federal roles in curriculum, testing and school choice are hotly contested, there is longstanding precedent and strong justification for Washington to create a major education manpower program. As in other key professions such as medicine, where the national government has long provided vital support for training and distributing doctors in shortage areas, the ability of schools to attract and retain well-trained teachers is often a function of forces beyond their boundaries. But without well-qualified teachers for schools with the neediest students, it will be impossible for them to make the progress on achievement in reading and mathematics that NCLB demands.

In that case, we would continue the historic pattern of failed federal education programs, in which low-income, disabled, language minority and other vulnerable students are taught by the least qualified teachers and untrained aides, rather than the skilled practitioners envisioned by the Elementary and Secondary Education Act and other national laws. The very purpose of these multibillion-dollar programs—to ensure equal education opportunity for the disadvantaged—has long been undermined by local inability or unwillingness to provide teachers capable of meeting the pupils' needs.

As the importance of well-qualified teachers for student achievement has become increasingly clear, this source of inequality has become more and more difficult to justify or ignore. On both equity and adequacy grounds, qualified teachers comprise a critical national resource that requires federal investment and cross-state coordination as well as other state and local action. No Child Left Behind provides a standard for equitable access to teacher quality that is both reasonable and feasible. Meeting this goal, however, calls for a new vision of the teacher labor market and development of a national teacher supply policy.

Understanding the Problems

To make headway on this agenda, it is essential to alter popular misunderstandings about teacher issues. For example:

- **The hiring of unqualified teachers is generally a result of distributional inequities, rather than overall shortages of qualified individuals.** Contrary to what some believe, the United States does not face an overall shortage of qualified teachers. While some schools **have dozens of qualified applicants for each position**, others—mostly those with poor and minority pupils—suffer from shortfalls, a mismatch that stems from an array of factors. They range from disparities in pay and working conditions, interstate barriers to teacher mobility and inadequate recruitment incentives to bureaucratic hiring systems that discourage qualified applicants, transfer policies that can slow hiring and allocate staff inequitably, and financial incentives to hire cheaper, less qualified teachers.

- **Retaining teachers is a far larger problem than training new ones—and a key to solving teacher "shortages."** In the years ahead, the chief problem will not be producing more new teachers, as many seem to believe. The main problem is an exodus of new teachers from the profession, with more than 30% leaving within five years. This, too, chiefly hurts low-income schools, which suffer from turnover rates as much as 50% higher than affluent schools (Ingersoll, 2001, p. 516). Such churning, which results in a constant influx of inexperienced teachers, is caused largely by insufficient preparation and support of new teachers, poor working conditions and uncompetitive salaries.

- **While the nation actually produces far more new teachers than it needs, some specific teaching fields do experience shortages.** These include teachers for children with disabilities and those with limited English proficiency as well as teachers of mathematics and physical science, two of the three subjects in which NCLB mandates student exams. Increasing supply in the few fields with shortfalls requires both targeted recruitment and helping preparatory institutions expand programs to meet select national needs.

To address these problems, we need to recognize that while teacher supply and demand historically have been local affairs, states and districts alone have been unable to solve these problems. Teacher issues increasingly are national in origin and consequences. While we should be mindful of the vital roles and prerogatives of states and localities, they need to be supported by appropriate national programs. These programs, we argue, should be modeled in good measure on U.S. medical manpower efforts, which have long supplied doctors to high-need communities and eased shortages in specific health fields. Similarly, teacher supply policy should help induce well-prepared teachers into districts that sorely need them—and enable them to succeed and stay there—while relieving shortages in fields like special education, math and physical science. It also should help stem departures of new teachers, which cost the nation billions of dollars a year. Indeed, the cost of the new programs could be entirely sustained by savings incurred by reducing teacher turnover.

The Alternative: Lowering Teacher Standards

The alternative to such policies is to lower standards for teacher knowledge and skills, through either continued emergency hiring or "quick-fix" programs that send people into difficult classrooms with little training in how to teach or deal with children. This has been the usual answer to teacher shortages, with unhappy results over the better part of a century. There are, fortunately, a growing number of new and rigorous alternate-certification programs based on careful selection, purposeful preparation, and intensive mentoring and practice teaching that are successful in preparing mid- career recruits from other fields. There is evidence that graduates of such programs feel confident about their teaching, are viewed as successful with children, and intend to stay in teaching (e.g. National Commission on Teaching and America's Future [NCTAF], 1996, p. 93; Miller, McKenna, & McKenna, 1998; Darling-Hammond, Kirby, & Hudson, 1989). We endorse these approaches.

However, we believe the evidence is clear that shortcut versions—those providing little training and meager support for new teachers—fail to prepare teachers to succeed or to stay, thus adding to the revolving door of ill-prepared individuals who cycle through the classrooms of disadvantaged schools, wasting district resources and valuable learning time for their students. Unfortunately, as some states develop plans to implement NCLB, they are including entrants into these programs (even before they have completed their modest training) in their definitions of "highly qualified" teachers.

The evidence to date provides cause for concern about this approach. For example, alternate-route teachers whose training lasts just weeks before they take over classes quit the field at high rates. Recent studies have documented such outcomes for recruits from the Massachusetts MINT program, nearly half of whom had left teaching within three years (Fowler, 2002) and the Teach for America program, an average of 80% of whom had left their teaching jobs in Houston, Texas, after two years (Raymond, Fletcher, & Luque, 2001). Analyses of national data show that individuals who enter teaching without student teaching (which these programs generally omit) leave teaching at rates twice as high as those who have had such practice teaching (Henke et al., 2000; NCTAF, 2003). Those who enter teaching without preparation in key areas such as instructional methods, child development and learning theory also leave at rates at least double those who have had such training (NCTAF, 2003, p. 84).

It is not hard to fathom why such teachers swiftly disappear. A former investment banking analyst, for example, tells of the "grim" circumstances she faced in a New York City elementary school, scarcely trained, unsupported, and realizing that "a strong academic background and years in an office are not preparation for teaching." Enthusiasm does not compensate for inexperience, she found, and teacher turnover is "so high that a school's 'veteran' teachers have frequently been around only three years, which makes it hard for new teachers to find experienced mentors." She quit after a year, part of the problem, not the solution. (Mehlman, 2002).

Despite this, the push to lower teacher standards, especially through quick-fix programs or back-door entry paths that skirt preparation, has strong adherents. These include some with influence in the U.S. Department of Education, as evidenced by the Secretary of Education's report to Congress on teacher quality. Called *Meeting the Highly Qualified Teachers Challenge* (U.S. Department of Education, 2002), the report is highly critical of teacher education, viewing certification requirements (Note 2) as a "broken system" and urging that attendance at schools of education, coursework in education and student teaching become "optional" (p. 19). By contrast, it regards alternate-route programs—especially those that eliminate most education coursework, student teaching and "other bureaucratic hurdles"—as the model option, getting teachers into classrooms on what it calls a "fast-track" basis. The report's prescription is for states to redefine teacher certification to stress content knowledge and verbal ability and to de-emphasize knowledge of how to instruct, assess, motivate or manage pupils.

The problem is not only that the report ignores and misrepresents research evidence, as has been documented in detail elsewhere. (Note 3) It is also that, together with other signals from Washington, it raises questions about how the Department of Education will enforce the requirement for all teachers to be highly qualified by the end of the 2005-2006 school year. "Highly qualified," according to NCLB, means that all teachers "must be fully licensed or certified by the state and must not have had any certification or licensure requirements waived on an emergency, temporary, or provisional basis." Teachers also must demonstrate subject matter competence (Title IX, Part A, Sec. 9101).

Now, however, the department appears to be signaling that states can comply in ways that dilute or undercut the law's standard. The statute permits "highly qualified" teachers to obtain full certification through traditional or alternative routes. However, the final regulations indicate that the department will accept state plans that designate as "highly qualified" those who have simply *enrolled* in alternative-certification programs, even if they have not completed them, demonstrated an ability to teach, or met the state's standards for a professional license. Such teachers may "assume the functions of a teacher" for up to three years without having received full certification and be considered "highly qualified." (Note 4) The department's comments on the final regulations make a point of noting that teachers in alternative routes to certification are to be considered an exception to the requirement that "highly qualified" teachers may not have had certification requirements "waived on an emergency, provisional, or temporary basis." (Note 5) The comments further suggest that "these alternative routes can also serve as models for the certification system as a whole." (Note 6)

Some states are proposing to meet NCLB requirements by lowering certification standards even further. For example, bills introduced in the 2002-2003 legislative sessions in Texas, Florida and California would allow candidates who have no preparation to teach to be certified so long as they have a bachelor's degree and pass a state test. In pressing for the Texas bill, (Note 7) the state comptroller argued that Texas should eliminate teacher education entirely from certification requirements, citing as her primary supporting evidence the Secretary of Education's report to Congress and speeches at a conference sponsored by the department (Strayhorn, 2003). The department, moreover, has signaled that it would welcome this further lowering of the bar on teacher standards.

Such interpretations of NCLB involve a sleight of hand on teacher qualifications. If certification requirements are redesigned to require less stringent standards than at present, meeting such standards will be an even poorer guarantee of teacher quality than what already exists. If some traditional teacher education programs have their flaws, essentially unregulated alternate-route programs lie almost completely beyond careful scrutiny. At this juncture in our history, encouraging the proliferation of untested alternatives raises the specter of a legally sanctioned, two-tiered staffing system. Schools that cannot afford competitive salaries, that cannot provide attractive working conditions, and that educate the most needy students will be staffed via untested alternate programs, while more advantaged schools will continue to recruit teachers with extended professional education. This certainly is not the intent of NCLB, but it could well be the result.

As we describe below, there is no research support for this approach. There *is* evidence, however, that it would reduce teacher effectiveness and contribute to teacher attrition. The chief victims would be the most vulnerable children in the hardest-to-staff schools, where underprepared teachers commonly work during their initial teaching years, before they meet licensing standards or leave the profession. This would extend the historic pattern of shortchanging disadvantaged students, even as evidence mounts that teacher quality is critical to student achievement. To cite just one of many studies, a 1991 analysis of 900 Texas school districts (Ferguson 1991) found that combined measures of teacher expertise—scores on a licensing examination, master's degrees and experience—accounted for more of the interdistrict difference in students' reading and mathematics achievement in grades 1 through 11 than any other factor, including students' family income. The effects were so strong and the variations in teacher quality so great that after controlling for socioeconomic status, the large disparities in achievement between black and white students were almost entirely accounted for by differences in teacher qualifications.

On the central importance of teachers there is, in fact, little disagreement, even among advocates for eased entry requirements. For example, the Thomas B. Fordham Foundation states, "The research shows that great teachers are the most important ingredients in any school. Smart, caring teachers can help their students overcome background problems like poverty and limited English proficiency." (Note 8) However, putting teachers with less preparation in classrooms for the neediest children will not provide equal opportunity or an adequate education. The far better strategy, we believe, is to craft a national teacher supply policy to ensure that well-prepared teachers are available to high-need districts, to produce more teachers in shortage fields, and to stem high teacher attrition rates. Even with such a system, of course, most decisions on teachers would remain the domain of state and local school officials, some of whom, as we shall see, have made important strides toward filling their classrooms with high-quality teachers—in part by doing exactly the opposite of what advocates for shortcuts recommend.

A Compelling State Interest

Those urging few certification requirements want to shift more decisions away from the states and to the local level. But states have a compelling interest in setting meaningful teacher standards. Murnane and colleagues (1991) note, for example, that traditional economic assumptions about consumer competence, priorities, knowledge, and information do not always hold with respect to teacher hiring, that "...some local districts (the purchasers of teachers' services) are underfunded, incompetent, or have priorities that the state finds unacceptable" (p. 94). If poor information were the only problem, then states could concentrate on requiring tests and other measures of the "right stuff," however defined. Local districts could then select based on scores and other information. However, if some local districts are likely to hire teachers whom the state finds unacceptable, then simple information alone will not solve this problem. The consequences of poor choices are not only local:

States are concerned because equal opportunity is threatened when incompetent teachers are hired, and the costs of inadequate education are borne not only by the children themselves, but also by the larger society. Dimensions of these costs include a lower rate of economic growth, higher incidence of welfare, greater crime rates, and higher unemployment rates (p. 95).

Economist Henry Levin (1980) makes a similar point:

> [T]he facts that we expect the schools to provide benefits to society that go beyond the sum of those conferred upon individual students, that it is difficult for many students and their parents to judge certain aspects of teacher proficiency, and that teachers cannot be instantaneously dismissed, mean that somehow the state must be concerned about the quality of teaching. It cannot be left only to the individual judgments of students and their parents or the educational administrators who are vested with managing the schools in behalf of society. The purpose of certification of teachers and accreditation of the programs in which they received their training is to provide information on whether teachers possess the minimum proficiencies that are required from the teaching function (p. 7).

Without strong, meaningful, and well-enforced certification requirements, not only will districts lack important information about candidates, but parents also will lack important safeguards regarding those entrusted with their children. In addition, states will lack the policy tools needed to encourage improvements in training and to equalize access to the key educational resource of well-prepared teachers.

To demonstrate why combining a national teacher supply program with state and local reform is the wiser way to meet the "highly qualified" teacher challenge, we examine the evidence on five issues:

- The kinds of teacher preparation that make a difference for student achievement.
- The evidence on alternative routes to certification.
- The current workings of the teacher labor market.
- The factors influencing teacher distribution.
- The steps some states and districts have been taking to ensure teacher quality.

We then turn to the elements of a national teacher supply policy for education.

I. What Preparation Makes a Difference in Student Learning?

There is wide agreement on some teacher attributes that appear to be related to teacher effectiveness and student learning. For example, virtually everyone acknowledges the importance of teachers' verbal ability and knowledge in the subjects taught. Those qualities, along with a liberal arts grounding, are at the heart of most state certification processes, which began requiring tests and coursework to assure competence in these areas in the early 1980s. These qualities are also central to the National Board for Professional Teaching Standards' voluntary certification process and other efforts to strengthen teacher education and professional development. The fact that alternative-certification advocates focus intently on such skills can only be welcomed. The problem is that these advocates very nearly stop there, as if little else mattered. Common sense and research evidence, however, tell us otherwise.

The Importance of Knowing How to Teach

Research shows that beyond verbal skills, subject matter knowledge and academic ability, teachers' professional knowledge and experience also make an important difference in student learning. Many other characteristics also matter for good teaching—enthusiasm, flexibility, perseverance, concern for children—and many specific teaching practices make a difference for learning (see e.g., Good & Brophy, 1995). The evidence suggests, in fact, that the strongest guarantee of teacher effectiveness is a combination of *all* these elements. (For reviews, see Darling-Hammond, 2000a; Wilson, Floden, & Ferrini-Mundy, 2001). It is this combination that most licensure processes seek to encourage, through requirements for courses, tests, student teaching and the demonstration of specific proficiencies.

Much of the research debate about what factors matter is due to the fact that few large-scale databases allow a comprehensive set of high- quality measures to be examined at once. Estimates of the relationships between particular teacher characteristics and student learning vary from study to study, depending on what factors are examined and when and where the study was conducted. Moreover, many variables that reflect teacher quality are highly correlated with one another. For example, teachers' education levels typically are correlated with age, experience and general academic ability. Similarly, licensure status is often correlated with academic skills, content background, education training and experience.

Studies linking teacher scores on tests of academic ability to student achievement (e.g. Coleman, et al., 1966; Ferguson & Ladd, 1996; Hanushek, 1992, 1996) have led some analysts to suggest that general academic or verbal ability are the primary measurable predictors of teacher quality. However, these studies typically have lacked other measures of teachers' preparation (for discussions, see Murnane, 1983; Wayne & Youngs, in press). When studies have looked directly at teachers' knowledge of both subject matter and how to teach, they have found that knowing how to teach also has strong effects on student achievement. Indeed, such studies show that knowledge of teaching is as important as knowledge of content (Begle, 1979; Monk, 1994; Wenglinsky, 2000).

For example, based on national survey data for 2,829 students, Monk (1994) found, not surprisingly, that teachers' content preparation, as measured by coursework in the subject field, was often positively related to student achievement in math and science. But courses in such subjects as methods of teaching math or science also had a positive effect on student learning at each grade level in both fields. For math, in fact, these teaching-method courses sometimes had "more powerful effects than additional preparation in the content area" (p. 142). Monk concluded that "a good grasp of one's subject area is a necessary but not a sufficient condition for effective teaching" (p. 142).

Wenglinsky (2002) looked at how math and science achievement levels of more than 7,000 8th graders on the 1996 National Assessment of Educational Progress (NAEP) were related to measures of teaching quality, teacher characteristics and student social class background. He found that student achievement was influenced by both teacher content background (such as a major or minor in math or math education) *and* teacher education or professional development coursework, particularly in how to work with diverse student populations (including limited-English-proficient students and students with special needs). Measures of teaching practices, which had the strongest effects on achievement, were related to teachers' training: Students performed better when teachers provided hands-on learning opportunities and focused on higher-order thinking skills. These practices were, in turn, related to training they had received in developing thinking skills, developing laboratory skills and having students work with real-world problems. The cumulative effect of the combined teacher quality measures, in fact, outweighed the effect of socioeconomic background on student achievement.

Teacher Certification and Student Learning

Since teacher certification or licensure has come in for criticism, we should look more closely at this factor. Although some analysts view licensure—or the teaching preparation that has typically been one of its major components—as unnecessary, the preponderance of evidence indicates that it, too, is associated with teacher effectiveness. Indeed, studies using national and state data sets have shown significant links between teacher education and licensure measures (including education coursework, credential status and scores on licensure tests) and student achievement. These relationships have been found at the level of the individual teacher (e.g., Goldhaber & Brewer, 2000; Hawk, Coble, & Swanson, 1985; Monk, 1994); the school (Betts, Reuben, & Danenberg, 2000; Fetler, 1999; Fuller, 1998, 2000; Goe, 2002); the school district (Ferguson, 1991; Strauss & Sawyer, 1986), and the state (Darling-Hammond, 2000a). The multi- level findings reinforce the inferences that might be drawn from any single study.

Goldhaber and Brewer (2000) concluded, for example, that the effects of teachers' certification on student achievement exceed those of a content major in the field, suggesting that what licensed teachers learn in the pedagogical portion of their training adds to what they gain from a strong subject matter background:

> [We] find that the type (standard, emergency, etc.) of certification a teacher holds is an important determinant of student outcomes. In mathematics, we find the students of teachers who are either not certified in their subject...or hold a private school certification do less well than students whose teachers hold a standard, probationary, or emergency certification in math. *Roughly speaking, having a teacher with a standard certification in mathematics rather than a private school certification or a certification out of subject results in at least a 1.3 point increase in the mathematics test. This is equivalent to about 10% of the standard deviation on the 12th grade test, a little more than the impact of having a teacher with a BA and MA in mathematics* (emphasis added). Though the effects are not as strong in magnitude or statistical significance, the pattern of results in science mimics that in mathematics (p. 139).

In this study, beginning teachers on probationary certificates (those who were fully prepared and completing their initial 2- to 3- year probationary period) from states with more rigorous certification exam requirements had positive effects on student achievement, suggesting the value of recent reforms to strengthen certification. (Note 9)

Similarly, a number of studies from states with large numbers of underprepared teachers have found strong effects of certification on student achievement. California is a case in point. There, three recent school-level studies found significant negative relationships between the percentage of teachers on emergency permits and student scores on state exams (Betts, Rueben, & Dannenberg, 2000; Fetler, 1999; Goe, 2002). Similarly, Fuller (1998, 2000) found that students in Texas schools with smaller proportions of certified teachers were significantly less likely to pass the Texas Assessment of Academic Skills (TAAS), after controlling for students' socioeconomic status and teacher experience.

This and other evidence suggests that it is a mistake to believe that one or two characteristics of teachers can explain their effects on student achievement. The message from the research is that multiple factors are involved and that teachers with a combination of attributes—knowing how to instruct, motivate, manage and assess diverse students, strong verbal ability, sound subject matter, and knowledge of effective methods for teaching that subject matter—hold the greatest promise for producing student learning. Those aspects of preparation that enable teachers to teach students with the greatest educational needs are, of course, most needed for teachers who will work with such children, a point that advocates of reduced standards for teachers in hard-to-staff schools (which serve these children) seem to miss. States and local districts should be pursuing fully prepared teachers, especially for the neediest students. They are the teachers whose training includes all of the attributes intended by the NCLB "highly qualified" definition.

II.The Evidence on Alternate Routes to Certification

The evidence on alternate-route programs is consistent with the research described above: In general, efforts that include a comprehensive program of education coursework and intensive mentoring have been found to produce more positive evaluations of candidate performance than models that forgo most of this coursework and supervised support.

Just as a quality distribution exists for conventional programs of teacher education, so there appears to be an even wider quality distribution for alternate programs (Darling-Hammond, Chung, & Frelow, 2002). At one end of the spectrum is a state alternative- certification program in New Hampshire that provides little structure or support. Candidates take "full responsibility for students prior to any preparation, and [have] three years to acquire 14 state- identified competencies through workshops or college courses" (Jelmberg, 1996, p.61). A study found that these alternate-route teachers were rated significantly lower than traditional teachers on instructional skills and instructional planning by their principals, and they rated their own preparation significantly lower than did traditionally certified teachers.

Some programs impart more systematic training and support. In a 1992 study of Connecticut's alternative-certification program—whose two-year training model provided "a significantly longer period of training than in any other alternate- route program" at the time (Bliss, 1992, p. 52)—supervisors gave mixed reviews of recruits' performance. Weaknesses were noted in relation to other teachers in terms of classroom management, but some strengths were found in teaching skills. A study of the Los Angeles Teacher Trainee Program, another two-year training model, also produced mixed results: University-trained English teachers were rated as more skillful than alternate-route (intern) teachers, while the levels of skill appeared more comparable but lower overall for math teachers from both groups (Stoddart, 1992).

In California, the Commission on Teacher Credentialing has worked to overcome shortcomings found in many local internship programs (McKibbin, 1998). A recent study of California State University teacher education graduates, however, found that those who prepared to teach after having entered teaching through emergency routes or internships felt less well prepared than those who had experienced a coherent program of pre-service preparation, and they also were perceived as less competent by their supervisors (California State University, 2002a; 2002b). A recent study by Stanford Research International echoed these concerns:

Principals reported that interns were less well prepared than fully credentialed recent hires in terms of their subject matter knowledge, their knowledge of instructional and assessment techniques, and their ability to teach basic skills to a diverse student population (Shields et al., 2001, p. 37).

The Dallas Schools' alternative-certification program provides summer training and then places recruits in mentored internships during the school year while they complete other coursework. In a study of this program, supervisors' perceptions of recruits were positive for the 54% who completed the intern year without dropping out or being held back due to "deficiencies" in one or more areas of performance (Lutz & Hutton, 1989). The study also reported data from an evaluation of the program by the Texas Education Agency (Mitchell, 1987), which surveyed principals, finding that:

The principals rated the [traditionally trained] beginning teachers as more knowledgeable than the AC interns on the eight program variables: reading, discipline management, classroom organization, planning, essential elements, ESL methodology, instructional techniques, and instructional models. The ratings of the AC interns on nine other areas of knowledge typically included in teacher preparation programs were slightly below average in seven areas compared with those of beginning teachers (Lutz & Hutton, 1989, p. 250).

Only two controlled studies of student achievement outcomes of alternate-route and traditionally trained teachers have been reported, again with mixed results. One, examining data from the Dallas program noted above, found that students of traditionally prepared teachers experienced significantly larger gains in language arts than those of alternate-route teachers (Gomez & Grobe, 1990). The other, using data from a well-designed program with strong pedagogical preparation and mentoring, found student outcomes comparable across the two groups (Miller, McKenna, & McKenna, 1998). This study focused on a university-sponsored program that provided 15 to 25 credit hours of coursework before interns entered classrooms. There they were intensively supervised and assisted by university personnel and school- based mentors while they completed additional coursework needed to meet full state licensure requirements. Because this design is so different from the many quick-entry, alternate-route programs, Miller, McKenna and McKenna (ibid) concluded that their studies

> . . . provide no solace for those who believe that anyone with a bachelor's degree can be placed

> in a classroom and expect to be equally successful as those having completed traditional education programs . . . The three studies reported here support carefully constructed AC programs with extensive mentoring components, post-graduation training, regular in-service classes, and ongoing university supervision (p.174).

One other program often cited in reference to alternative certification is Teach for America, although TFA is a recruiting program rather than an alternative-certification program. After controlling for teacher experience and school and classroom demographics, one study found that TFA recruits in Houston were about as effective as other inexperienced teachers in schools and classrooms serving high percentages of minority and low-income students, which is where most underqualified teachers in the district are placed (Raymond et al., 2001). In 1999-2000, the last year covered by the study sample, about 50% of Houston's new teachers were uncertified, and the researchers reported that 35% of new hires lacked even a bachelor's degree, so TFA teachers were compared to an extraordinarily ill-prepared group. Raymond and colleagues did not report how TFA teachers' outcomes compare to those of trained and certified teachers. However, a separate study in Arizona that examined this question found that students of TFA teachers did significantly less well than those of certified beginning teachers on math, reading and language arts tests (Laczko-Kerr & Berliner, 2002).

Ideally, we would like to know more about the effectiveness of different kinds of alternate-route programs. Although the research is not definitive (see Wilson, Floden, & Ferrini-Mundy, 2001 for one synthesis, SRI International, 2002, for another), most studies to date tend to support more extensive training over speeding recruits into classrooms with little preparation or support.

Given the evidence suggesting the importance of the preparation intended by NCLB, the question is whether it is possible for states to comply in the face of what appear to be substantial teacher shortages in some places? The evidence suggests that states can indeed comply—with targeted policies that better organize and more equitably distribute their own teaching force, supplemented with a national system that, among other things, works to correct the maldistribution of well-qualified teachers.

III. The Teacher Labor Market

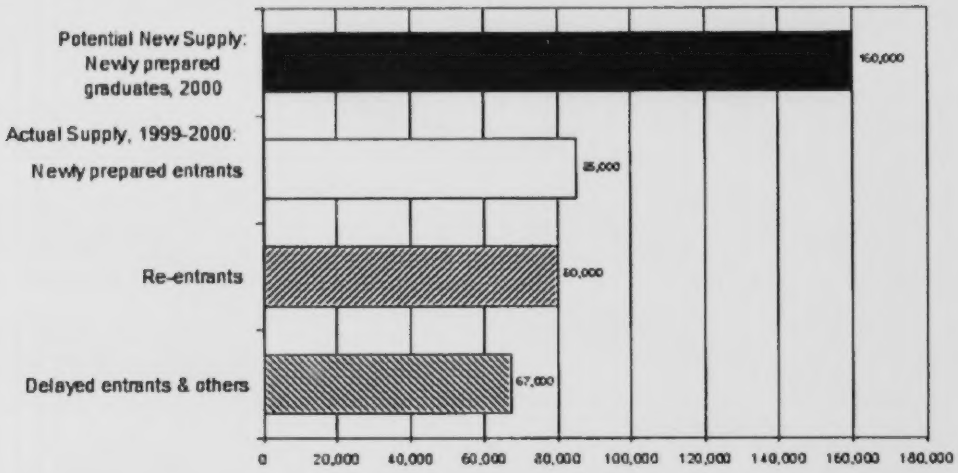
To understand how teachers become so inequitably distributed, we need to examine how teacher supply and demand operate, what causes teacher attrition, and why there are teacher shortages in particular fields. We will then look at the chief causes of the inequitable distributions that are the target of No Child Left Behind.

More Supply Than Demand. The nation currently is in the midst of a teacher hiring surge that began in the early 1990s. Annual demand recently has averaged about 230,000 teachers—demand that can easily be met with existing well-prepared teachers from our three main supply sources. Only one of these sources is newly prepared teachers, who generally constitute no more than half the teachers hired in a given year. (Note 10) In 1999, for example, when U.S. schools hired 232,000 teachers who had not taught the previous year, fewer than 40% (about 85,000) had graduated from college the year before. About 80,000 were from the second source—re-entrants from the reserve pool of former teachers (NCTAF, 2003). (Note 11) Of the remaining 67,000, most were from the third source—delayed entrants who had prepared to teach in college but who had taken time off to travel, study, work in another field or start a family. (Note 12)

In the aggregate, worries about preparing many more new teachers to meet demand are misplaced. As a nation, we produce many more new teachers than the 100,000 or fewer that are needed annually. In 2000, for example, the 603 institutions counted in the AACTE/NCATE joint data system—representing about half of all teacher training institutions and about three-quarters of teachers in training—reported 123,000 individuals who completed programs that led to initial teaching certification. So the newly prepared pool that year was well above 160,000, (Note 13) before counting those who entered teaching through alternative pathways that were not university-based. (Note 14) (see Figure 1). Overall, according to the U.S. Census Bureau, more than 6 million people in the nation held a bachelor's degree in education in 1993. This represented only a fraction of the credentialed teacher pool, since most teachers now enter teaching with a major in a disciplinary field plus a credential or master's degree in education. So excluding the 2.5 million active teachers at that time, more than 4 million people were prepared to teach but were not doing so.

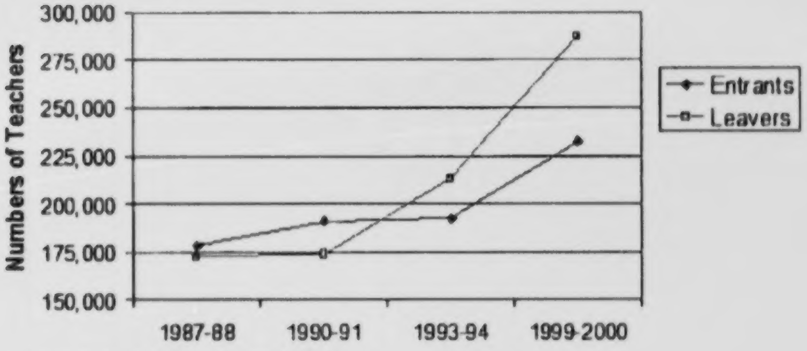
If we have no overall "shortage" of individuals prepared to teach, why are there so many unqualified teachers in some states and cities? What we do have is a maldistribution of teachers, with surpluses in some areas and shortfalls in others. In 2000, for example, there were surpluses of teachers in most fields in the Northwest, the Mid-Atlantic and much of the South but shortages in the far West, the Rocky Mountain States, and Alaska (American Association for Employment in Education, 2000). With slowed employment in other sectors of the economy during 2002 and teacher salary hikes in some places that had previously had hiring problems, newspapers across the country carried stories of shortages being resolved (see, e.g. Gormley, 2003; Zhao, 2002). In some growing areas, enrollment increases will likely continue to create hiring pressures, while enrollment declines promise to expand teacher surpluses elsewhere. By 2007, for example, enrollments are projected to climb by more than 20% in California and Nevada while shrinking in most parts of the Northeast and Midwest. But enrollment levels are not the central problem.

Figure 1 - Sources of Teacher Supply



The Exodus of Beginning Teachers. A much larger challenge than preparing new teachers is retaining existing teachers. Since the early 1990s, the annual outflow from teaching has surpassed the annual influx by increasingly large margins, straining the nation's hiring systems. While schools hired 232,000 teachers in 1999, for example, 287,000 teachers left the profession that year (see Figure 2). Retirements make up a small part of this attrition. Only 14% of teachers who left in 1994-1995 listed retirement as their primary reason (Ingersoll, 2001). More than half left to take other jobs and/or because they were dissatisfied with teaching. Especially for hard-to- staff schools, the largest exodus is by newer teachers who are dissatisfied with working conditions or have had insufficient preparation for what they face in classrooms (Ingersoll, 2001; Henke, et al., 2000).

Figure 2 - Trends in Teacher Entry and Attrition, 1987-2000



Source: Adapted from Ingersoll (2001).

The early exodus of teachers from the profession has been a longstanding problem. Studies indicate that as many as 20% of new teachers may leave teaching after three years and that closer to 30% quit after five years. (Note 15) Departure rates for individual schools and districts run higher, as they include both "movers," who leave one school or district for another, as well as "leavers," who exit the profession temporarily or permanently. Together, movers and leavers particularly affect schools serving poor and minority students. Teacher turnover is 50% higher in high-poverty schools than in more affluent ones (Ingersoll, 2001, p. 516), and new teachers in urban districts exit or transfer at higher rates than suburban counterparts (Hanushek, Kain, & Rivkin, 1999). In addition, teachers quit schools serving low-performing students at much higher rates than they quit successful schools (Hanushek, Kain & Rivkin, 1999, p. 15). As a result, these schools are often staffed disproportionately with inexperienced as well as ill-prepared teachers.

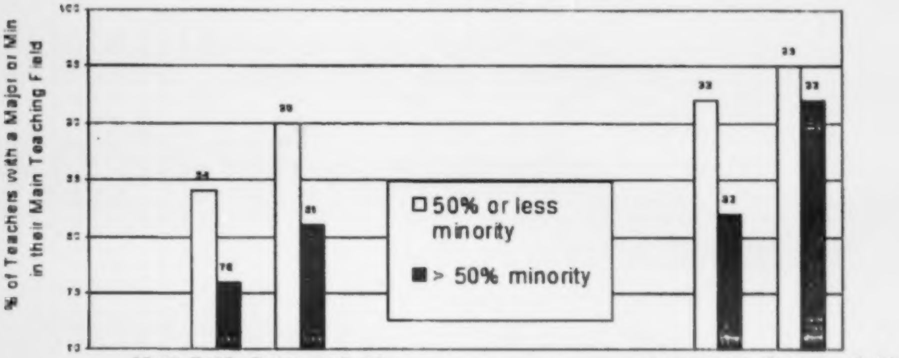
The costs of early departures from teaching are immense, as evidenced by a recent study in Texas that employed different models to estimate the costs of teacher turnover. Based on the state's current turnover rate of 15.5%, which includes more than 40% of beginning teachers quitting the field in their first three years, the study found that, "Texas is losing between $329 million and $2.1 billion per year, depending on the industry cost model that is used" (Benner, 2000, p. 2). This represents between $8,000 and $48,000 for each beginning teacher who leaves. The larger figure, truly a staggering number, stems from a model that includes separation costs, replacement or hiring costs, training costs, and learning-curve loss. Using even the lowest estimate for this one state, however, it is clear that early attrition from teaching costs the nation billions of dollars each year.

Such churn among novices also reduces overall education productivity, since teacher effectiveness rises sharply after the first few years in the classroom (Hanushek, Kain, & Rivkin, 1998; Kain & Singleton, 1996). It drains affected schools' financial and human resources. These schools, which typically can least afford it, must constantly pour money into recruitment and professional support for new teachers, many of them untrained, without reaping benefits from the investments. Other teachers, including the few who could serve as mentors, are stretched thin by the needs of their colleagues as well as their students (Shields et al., 2001). Scarce resources are wasted trying to re-teach the basics each year to teachers who arrive with few tools and leave before they become skilled (Carroll, Reichardt, & Guarino, 2000). Most important, the constant staff churn consigns a large share of children in high-turnover schools to a parade of relatively ineffective teachers.

Shortage Fields. While U.S. teacher supply is sufficient on the whole to meet demand, there are nonetheless longstanding shortages in particular fields. These result largely from more attractive earnings opportunities outside teaching. Mathematics and science teaching, for example, suffer larger wage disparities than those for English and social studies. Thus college graduates trained in mathematics and the sciences typically must forgo greater salaries in order to teach. Likewise, increased demand for special education and bilingual education teachers, and the skill sets that trained teachers in these fields possess, have produced shortfalls in many states and localities. (Note 16)

These shortages, again, particularly hurt disadvantaged students. This is not only because of pupils taught by unqualified special education and bilingual education teachers. It is also because less advantaged minority students disproportionately end up with unqualified teachers of science and math as well. In 1993-1994 only 8% of public school teachers in wealthier schools taught without a major or minor in their main academic assignment—compared with fully a third of teachers in high-poverty schools. Moreover, nearly 70% of those in poor, minority schools taught without at least a minor in their secondary field (National Center for Education Statistics, 1997). In 1998, the proportions of out-of-field math and science teachers, though somewhat lower, were still much higher in low-income, minority and urban schools (NCES, 2000) (see Figure 3).

Figure 3- Disparities in Access to Qualified Math and Science Teachers, 1998



Source NCES, 2000

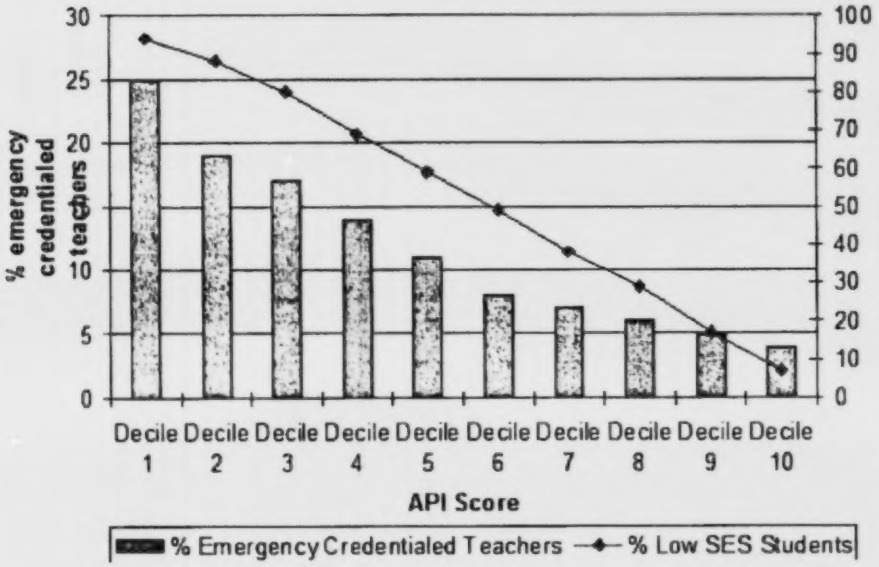
The Children Who Suffer Most. With all of these problems—whether the general maldistribution of teachers, the exodus of younger teachers from the profession, or shortages in special fields—the chief victims are disadvantaged students in big cities or poor rural areas. This heavily reflects the nation's inequitable funding of education. In most states, the wealthiest districts have revenues and expenditures per pupil that are two or three times those of the poorest districts (Educational Testing Service, 1991; Kozol, 1991). Poor rural districts typically spend the least, and urban districts serving students with multiple needs spend much less than surrounding suburbs, where students and families have far fewer challenges. These inequities translate into differentials in salaries and working conditions—resources that greatly affect teacher labor markets.

A recent report from the Education Trust (2002) found that, in many states, the quartile of districts with the highest child poverty rates receives less state and local funding per pupil than the most affluent quartile. The study indicated that, nationwide, this disparity decreased slightly between 1997 and 2000, a somewhat hopeful sign. (Note 17) Nevertheless, the disparities persist, and their effects are amplified by the needs students bring to school. A recent large-scale study of young children found that children's socioeconomic status (SES) is strongly related to cognitive skills at school entry. For example, the average cognitive scores of entering children in the highest SES group are 60% above the average scores of the lowest SES group (Lee & Burkham, 2002). As the study documents, low-SES children then begin kindergarten in systematically lower-quality schools than their more advantaged peers, no matter what measure of quality is used—qualified teachers, school resources, teacher attitudes, achievement or school conditions. From the outset of schooling, then, inequalities associated with family circumstances are multiplied by inequalities of education.

Those unequal opportunities then continue throughout the students' educations. In almost every field, central city schools with the largest numbers of disadvantaged children are much more likely than other schools to report unfilled teacher vacancies (NCES, 1997, Table A8.11). These schools are also far more likely than others to fill vacancies with unqualified teachers. The funding inequalities also lead to enlarged class sizes and lack of access to higher-level courses as well as to poorer teaching (Choy, et al., 1993).

California data provide a dramatic example of the maldistribution of qualified teachers and its effects. On the one hand, many California districts have little difficulty hiring qualified teachers. In 2000- 2001, for example, about 47% of districts (41% of schools) had fewer than 5% uncredentialed teachers, and about 25% hired no unqualified teachers at all (Shields, et al., 2001, p. 21-23). However, in another quarter of California schools, more than 20% of teachers were underqualified (i.e., lacking a preliminary or professional clear credential), and in some schools a majority of teachers lacked full certification. As Figure 4 shows, the presence of underqualified teachers is strongly related both to student socioeconomic status and to student achievement, with students who most need highly qualified teachers least likely to get them.

Figure 4
The Relationship between California Elementary School API Scores, Student Socioeconomic Status, and Teacher Qualifications, 2000



Across the nation, disparities in access to qualified teachers occur not only among districts but also among schools within districts. Among other things, recent studies show:

- Nonwhite, low-income and low-performing students, particularly in urban areas, are disproportionately taught by less qualified teachers (Hanushek, Kain, & Rivkin, 2001; Ingersoll, 2002; Jerald, 2002; Lankford, Loeb, & Wyckoff, 2002).
- Teachers most often transfer out of schools with poor, minority, and low-achieving students (Ingersoll, 2001; Lankford, Loeb, & Wyckoff, 2002; Hanushek, Kain, & Rivkin, 2001; Scafidi, Sjoquist, & Stinebrickner, 2002).
- School and district disparities in teacher qualifications persist over time and have worsened in the past 10 to 15 years as teacher demand and funding inequities both have increased (Jerald, 2002; Lankford, Loeb, & Wyckoff, 2002; NCES, 2002).

IV. What Factors Influence Teacher Distribution?

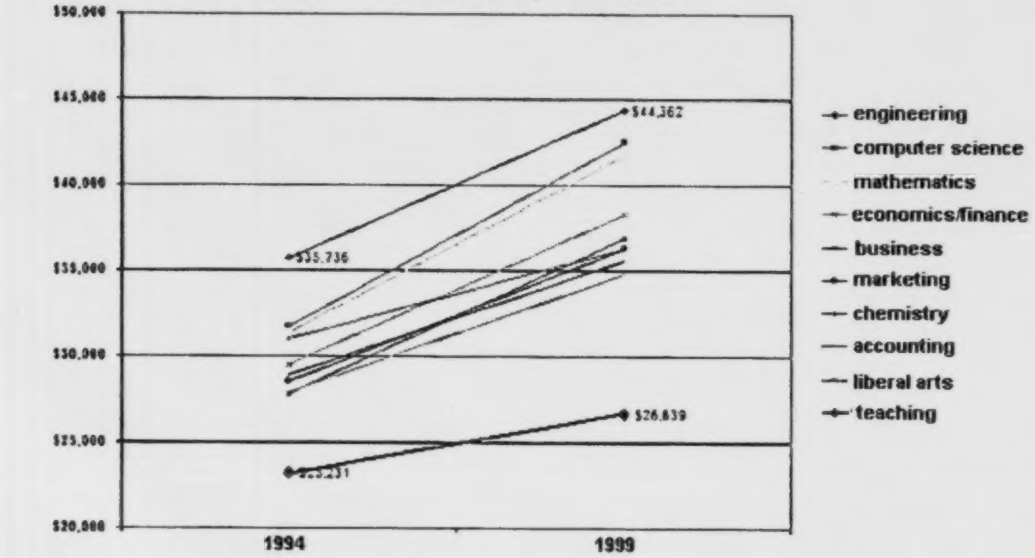
Researchers have examined what factors influence who teaches where and how long they stay. These include wages and benefits, "non- pecuniary" considerations such as working conditions and student characteristics, teacher preparation and district personnel policies (Lankford, Loeb, & Wyckoff, 2002, pp. 38-39). Disentangling these factors is essential to the evaluation of policy alternatives. If teachers generally prefer teaching white, middle-class, high-performing students, for example, that preference may be hard to influence. But if teachers object to working conditions that often attend teaching poor and minority children, those are potentially alterable. Many analysts (e.g., Ballou, 1996, Ballou & Podgursky, 1997; Wise, Darling- Hammond, & Berry, 1987) also contend that districts and schools often fail to hire the best candidates, at any given salary level, introducing inefficiencies into the labor market for teachers. So the joint preferences of individuals and organizations interact to determine who teaches and where they teach. A brief tour of this terrain suggests the kinds of policies needed.

The Draw of Home. The first feature of note is the longstanding tendency for many teachers to seek positions close to where they grew up or, to a lesser extent, went to college. As Boyd and colleagues (2003) note: "The importance of distance in teachers' preferences particularly challenges urban districts, which are net importers of teachers" (p.12). While teachers who grew up in cities often are inclined to teach in their hometowns, the number of urban recruits falls short of the number needed, requiring urban districts to seek teachers from elsewhere. If urban districts cannot offer compensating incentives, urban recruits are likely to be less qualified overall than those who teach in suburbs. The differential qualifications of teachers in disadvantaged urban schools appear to be at least as much a function of first-job placements as differential exits or transfers accounts. Geography, then, clearly plays a powerful role, a point to which we return in our policy recommendations.

Salaries. Even if teachers may be more altruistically motivated than many other workers, teaching must compete for talented college graduates in ways that include pay. On this score, although overall teacher demand can be met, there is reason for concern. Teacher pay not only is relatively low, but during the 1990s it also declined relative to other professional salaries (see Figure 5). Even after adjusting for the shorter work year in teaching, teachers earn 15% to 30% less than college graduates who enter other fields.

Today's troubled economy is temporarily offsetting these trends because of the relative stability of teaching compared with such hard-hit sectors as high technology. Thus in the Silicon Valley area, the flow of technology workers into math and science teaching recently has swelled, and reports indicate that applications are up elsewhere as well (Hayasaki, 2003). The profession needs to maximize this temporary opportunity, ensuring that enough new entrants, especially from high- need fields, receive sufficient training and support to succeed, adding to the long-term supply of high-quality teachers. Otherwise, demand from career-switchers may increase pressure for fast-track training, creating teachers who may soon become part of the exodus from the profession. It is important to recognize, moreover, that the economy's cycles are temporary, so before too long many career- switchers may return to more lucrative occupations if they do not find satisfying work in teaching. What happens with respect to school revenues, teacher salaries, and subsidies for decent training for these new entrants will determine whether schools can benefit from these trends.

**Figure 5 - Trends in Starting Salaries across Professions
United States, 1994-1999**



There is evidence that wages are at least as important to teachers in their decision to enter and quit the profession as they are to workers in other occupations (Baugh and Stone, 1982). Teachers are more likely to leave the field when they work in districts with lower pay and when their salaries are low compared to other wage opportunities (Brewer, 1996; Mont & Rees, 1996; Murnane, Singer & Willett, 1989; Theobald, 1990; Theobald & Gritz, 1996). These factors are strongest at the start of the teaching career (Hanushek, Kain & Rivkin, 1999; Gritz & Theobald, 1996) and for teachers in high- demand fields like math and science (Murnane and Olsen, 1990; Murnane, et al., 1991).

But do pay increases result in better educational results? To find out, some analysts have examined the relationship between changes in teacher salaries and student achievement. Based on a meta-analysis of about 60 production function studies, for example, Greenwald, Hedges and Laine (1996) found larger effects for student achievement associated with increased teacher salaries (as well as with teacher experience and education, which are rewarded in salary schedules) than for such other resources as reduced pupil-teacher ratios. Ferguson's (1991) analysis of student achievement in Texas also concluded that student gains were associated with the use of resources to purchase higher-quality teachers. In an analysis of hiring practices and salaries in California counties, Pogodzinski (2000) found that higher salaries appeared to attract better-prepared teachers. Finally, in a study looking across states from 1960 through 1990 and across districts in California from 1975 through 1995, Loeb and Page (2000) found that student educational attainment increased most in states and districts that increased teacher wages.

Studies confirm that salaries are widely disparate both within and across states—and that school systems serving large numbers of low-income and minority students often have lower salary levels than surrounding districts (Lankford, Loeb, & Wyckoff, 2002). Nationally, teachers in schools serving the largest concentrations of low-income students earn, at the top of the scale, salaries one-third less than those in higher-income schools (NCES, 1997), while they also face lower levels of resources, poorer working conditions, and the stresses of working with students and families who have an array of needs. Pogodzinski (2000) found that large differences in teachers' wages across schools districts within the same county are a significant factor in explaining the use of emergency permits and waivers.

Once teachers begin work, however, transfers to other schools often appear to be influenced only modestly by salaries and more by other factors (Loeb & Page, 2000). While one study found that teacher transfers tended to improve salary slightly (Hanushek, Kain, & Rivkin, 2001), another found that salary variation seemed to contribute little to teacher sorting among schools (Lankford, Loeb, & Wyckoff, 2002). We conclude, then, that teacher salaries are important in attracting individuals to teaching from the college-educated pool and in influencing early career behavior. They also have an effect on attrition. But other factors also matter to teachers' decisions about whether and where to continue teaching.

Working Conditions and Dissatisfaction. Surveys have long shown that working conditions play a large role in teacher decisions to change schools or leave the profession. Reasons for remaining in teaching or leaving are strongly associated with such matters as how teachers view administrative support, available education resources, teacher input into decisionmaking, and school climate (Darling-Hammond, 1997; Ingersoll, 2001, 2002). Moreover, there are large differences in the support teachers receive in affluent and poor schools. Teachers in more advantaged communities experience easier working conditions, including smaller class sizes and pupil loads, more materials and greater influence over school decisions (NCES, 1997, Table A 4.15). In 1994-1995, more than a quarter of all school leavers listed dissatisfaction with teaching as a reason for quitting, with those in high-poverty schools more than twice as likely to leave because of this than those in wealthier schools (Darling-Hammond, 1997).

A number of studies have found that teacher attrition appears related to student demographics, with teachers transferring out of high-minority and low-income schools (e.g., Carroll, Reichardt, & Guarino, 2000; Scafidi, Sjoquist, & Stinebrickner, 2002) or out of low-performing schools into better-performing ones (Hanushek, Kain, & Rivkin, 2001). Given the confluence of negative conditions in schools serving low-income and minority students, the question is whether these demographic variables can be disentangled from other non- pecuniary factors that are amenable to policy influences.

There is evidence that working conditions are an important independent cause of teacher attrition, beyond the student characteristics frequently associated with them. For example, a survey of California teachers (Harris, 2002) found that teachers in high- minority, low-income schools reported significantly worse working conditions, including poorer facilities, fewer textbooks and supplies, less administrative support and larger class sizes. Furthermore, the teachers were significantly more likely to say that they planned to leave a school soon if working conditions were poor. The relationship between teachers' plans to leave and schools' demographic characteristics was much smaller.

A multivariate analysis of these California data found that turnover problems at the school level are, in fact, influenced by student characteristics, but that demographic variables become much less significant when working conditions and salaries are considered. Working conditions—ranging from large class sizes and facilities problems to multi-track, year-round schedules and faculty ratings of teaching conditions—proved to be the strongest predictors of turnover problems, along with salaries (Loeb, Darling-Hammond, & Luczak, forthcoming). We believe that such conditions constitute a primary target for policies aimed at retaining qualified teachers in high-need schools.

Finally, a new aspect of working conditions that affects teacher retention may be traced to unexpected consequences of the new accountability. In many states today, schools that fail to meet performance standards on state assessments are being targeted for special attention, often associated with new labels. Low-performing schools frequently are identified in the local press and may be subject to sanctions and interventions. Such targeting can be valuable in identifying schools that most need more help, but it can also stigmatize such schools, affecting staff morale and leading to a teacher exodus. Evidence of such effects is beginning to emerge. A Florida report described teachers leaving schools rated "D" or "F" in "droves" (DeVise, 1999). A

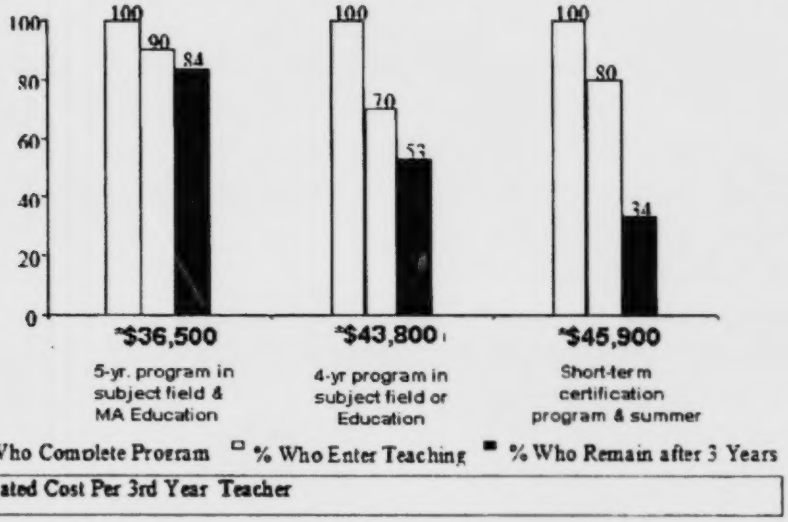
North Carolina study found "failing" schools lagging behind others in their ability to attract more highly qualified teachers, a trend researchers attribute to the accountability system (Clotfelter et al., 2003). In the California study noted above, teachers rated more negatively than any other working condition the state tests they are required to administer. This was a component of the measure that significantly predicted turnover (Loeb, Darling-Hammond, & Luczak, 2003).

Teacher Preparation and Support. A factor often overlooked in economic analyses is the effect of preparation on teacher retention. A growing body of evidence indicates that attrition is unusually high for those with little initial preparation. A recent NCES study found, for example, that 49% of uncertified entrants left the profession within five years, more than triple the 14% of certified entrants who did so (Henke, et al., 2000). This report and an analysis of another NCES data base both showed attrition rates for new teachers who lacked student teaching at rates double those of those who had had student teaching (NCTAF, 2003).

In California, the state standards board has found that 35% to 40% of emergency permit teachers leave the profession within a year (Darling-Hammond, in press; Tyson, Hawley, & McKibbin, 2000, p. 3). National data from the Recent College Graduates Survey indicate that about two-thirds of novices who enter without teacher education (neither certified nor eligible for certification) leave teaching within their first year (Grey, et al., 1993). As noted previously, moreover, studies of entry paths to teaching that offer only a few weeks of training before assumption of full teaching responsibilities have also found high attrition rates.

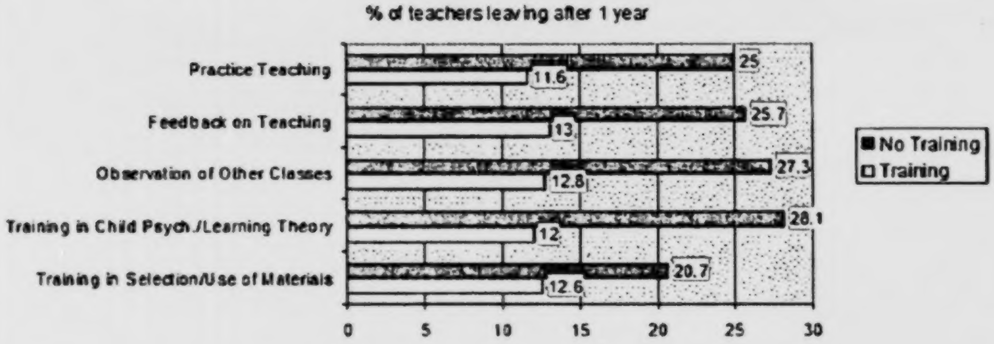
Conversely, accumulating evidence indicates that better-prepared teachers stay longer. For example, a longitudinal study of 11 institutions found that teachers who complete redesigned 5-year teacher education programs enter and stay in teaching at much higher rates than 4-year teacher education graduates from the same campuses (Andrew & Schwab, 1995). The 5-year programs allow a major in a disciplinary field, intensive training for teaching and long-term student teaching. In addition, both 4- and 5-year teacher education graduates enter and stay at higher rates than teachers hired through alternatives that offer only a few weeks of training before recruits are left on their own in classrooms (Darling-Hammond, 2000b). These differences are so large that, considering the costs to states, universities and school districts of preparing, recruiting, inducting and replacing teachers due to attrition, the cost of preparing a career teacher through a 5- year program is actually far less than that of preparing larger numbers, many of whom leave, through short-term routes (see Figure 6). Graduates of 5-year programs also report higher levels of satisfaction with their preparation and receive higher ratings from principals and colleagues.

Figure 6 Average Retention for Different Pathways into Teaching



Similarly, Schools and Staffing Survey (SASS) data for 1999-2000 show big differences in plans to stay in teaching between first-year teachers who felt well prepared and those who felt poorly prepared. On such items as preparation in planning lessons, using a range of instructional methods and assessing students, two-thirds of those reporting strong preparation intended to stay, compared to only one- third of those reporting weak preparation. The differentials hold true for actual attrition as well. Analyses of SASS Teacher Follow-up data show that new recruits who had training in such aspects of teaching as selecting instructional materials, child psychology and learning theory, who had practice teaching experience and who received feedback on their teaching left the profession at rates half as great as those who did not have such preparation (NCTAF, 2003) (see Figure 7). Similarly, a survey of 3,000 beginning teachers in New York City found that recruits who felt better prepared were more inclined to stay in teaching, to feel effective, and to say they would enter through the same program or pathway again. Graduates of teacher education programs felt significantly better prepared and more effective than those entering through alternative routes or with no training (Darling- Hammond, Chung, and Frelow, 2002).

Figure 7 - Teacher Preparation Reduces Attrition of 1st Year Teachers



The effects of strong initial preparation are likely to be enhanced by equally strong induction and mentoring in the early teaching years. School districts such as Cincinnati, Columbus and Toledo, Ohio, and Rochester, New York, have reduced beginning-teacher attrition rates by more than two-thirds by providing expert mentors with release time to coach beginners in their first year (NCTAF, 1996). These young teachers not only stay in the profession at higher rates, but they also become competent more quickly than those who learn by trial and error.

States increasingly are requiring induction programs, some with strong results. Unfortunately, quality can decline as programs expand. In an assessment of one of the oldest, California's Beginning Teacher Support and Assessment (BTSA) Program, for example, early pilots with carefully designed mentoring systems found rates of new- teacher retention exceeding 90% in the first two to three years on the job. However, as the program scaled up with more uneven implementation across the state, a later study reported that only 47% of BTSA participants had received classroom visits from their support provider at least monthly, and only 16% of novice teachers participating in other induction programs had received such visits. Often, districts provided orientation sessions and workshops rather than on-site mentoring, the most powerful component of induction programs (Shields, et al., 2001, p. 101). While state induction programs for beginning teachers rose from seven in 1996-97 to 33 in 2002, only 22 states fund the programs, and many do not require regular, on-site coaching (NCTAF, 2003). To reap the gains that well-designed programs have realized, state-mandated induction programs must include real support and follow- through.

Particularly in hard-to-staff schools, then, policies encouraging strong initial teacher education are warranted, along with strong induction and continuing support. Initial preparation cannot overcome poor working conditions and inadequate support, but it can launch teachers successfully, reducing the odds that they will leave teaching altogether.

Personnel Management. Finally, how districts and schools—within the constraints of state policies and collective bargaining agreements—recruit, hire, assign, support and manage transfers of teachers plays a large role in determining shortages. Studies in locales ranging from large cities to small rural districts make clear how local management preferences and practices shape who teaches in which schools—and how such preferences can systematically enhance or undermine both efficiency and effectiveness.

Some states, for example, enforce redundant requirements for fully qualified and credentialed candidates from other states, making it difficult for them to enter the local teaching force. (Note 18) Additional barriers include late budget decisions by state and local government, teacher transfer provisions that push new hiring decisions into August or September, lack of pension portability across states and loss of salary credit for teachers who move. Nor does the list stop there. For example, most districts have salary caps for experienced candidates. As a result, some highly desirable teachers must take pay cuts if they want to teach in new schools where they have moved. Changing professions can look like a better option in those circumstances. Likewise, few districts reimburse travel and moving expenses, yet another barrier to mobility in the teacher labor market.

Atop all of this, many districts do not hire the best applicants because of inadequate information systems or antiquated and cumbersome procedures that discourage or lose candidates in seas of paperwork (Wise, Darling-Hammond, & Berry, 1987). For example, before its recent overhaul, the 62-step hiring process in Fairfax County, Virginia, mirrored those of many other large districts that attract a surplus of qualified applicants but cannot find an efficient way to hire them (NCTAF, 1996). A process that takes months and features long lines and delays can discourage all but the most persistent.

In districts with high demand relative to supply, late hiring and disorganized hiring processes can undermine the recruitment of qualified teachers. In one recent study, conducted in four states, researchers found that one-third of a sample of new, young teachers were hired after the school year had already started; only 23% had any sort of reduced load; 56% received no extra assistance; and 43% went through the entire first year with no observations from a mentor or more experienced teacher (HGSE, 2003, April). In another study, nearly 50% of newly hired California teachers were hired after August 1, and 25% were hired after the start of the school year (Shields, et al., 1999). Teachers in schools with large numbers of underprepared teachers were significantly less likely to report that they had been actively recruited or assisted in the hiring process and more likely to report that the hiring process had been slow and filled with obstacles (Shields, et al., 2001, p. 84). The California State Fiscal and Crisis Management Team reports hiring and screening procedures that are erratic and fraught with glitches, application processes that are not automated or well-coordinated, applicants and vacancies that are not tracked, and recruitment that is disorganized in districts that hire large numbers of underqualified teachers (Darling-Hammond, in press).

Various studies have uncovered still more reasons for district hiring of unqualified teachers. These include patronage, a desire to save money on salaries by hiring low-cost recruits over better- qualified ones, and beliefs that more qualified teachers are more likely to leave and less likely to take orders (Pflaum & Abramson, 1990; Schlechty, 1990; Wise, Darling-Hammond, & Berry, 1987). Testimony before the California Assembly Select Committee on Low Performing Schools (2001) pointed to the prevalence of such concerns:

> [I]n some situations districts hire emergency permit holders because [they] can be paid less; need not initially be provided with benefits; cannot be placed on a tenure track; can be dismissed easily; and need not be provided with systematic support and assistance... (p. 5).

Yet other influences on the assignment of teachers may operate at the school level. In schools serving advantaged families, parents will tolerate less mediocrity in teaching and are more likely to exert pressure to hire and retain well-qualified teachers. At the classroom level, some parents pressure administrators to obtain or avoid certain teachers for their children. Responding to such informal pressures may systematically alter the availability of effective teachers for students who lack vocal and knowledgeable parent advocates. Such informal, "micro-level" processes are likely to operate unless countervailing tendencies are present (see Bridges, 1990, Clotfelter, et al., 2003).

Finally, in many states collective bargaining agreements influence the effective deployment of teachers. In particular, contract provisions that regulate transfers among schools by seniority often mean that hard-to-staff schools systematically lose experienced teachers. Turnover in such schools is high, with a steady influx of young, inexperienced teachers who often are ill supported by mentor or induction programs. In some locales, progressive labor-management relations have resulted in bargaining agreements that create more equitable staffing patterns, but these are the exceptions.

Several critical points emerge from this thicket of issues. First, incentives that influence teacher entry and mobility often fail to support an equitable distribution of teachers across districts, schools and classrooms. Salaries and working conditions are unequal, and they fail to provide compensating inducements in support of hard-to-staff schools. Second, teacher preferences and school system behaviors influence teacher distribution. Many states and districts manage hiring inefficiently for reasons ranging from fiscal conditions to management procedures, contract provisions and parent pressures. Taken together, these factors create a maldistribution of teachers that is systemic in nature and that will require coordinated responses across the levels of government and education to solve. As we discuss in the next section, some locales have begun to develop policies and practices that make genuine headway on these problems. These and other exemplars suggest how policies can be developed that directly address the sources of longstanding disparities.

V. Lessons from State and District Experiences

In this section, we describe examples of both states and local school districts that have fashioned successful strategies for strengthening their teaching forces. These approaches inform our recommendations at the end of this paper.

A. State Approaches

Beginning in the 1980s, Connecticut and North Carolina enacted some of the nation's most ambitious efforts to improve teaching. On the heels of these efforts, these states, which serve sizable numbers of low-income and minority students, (Note 19) registered striking gains in overall student learning and narrowed achievement gaps between advantaged and disadvantaged pupils. During the 1990s, for example, North Carolina posted the largest student achievement gains of any state in math and sizable advances in reading, putting it well above the national average in 4th grade reading and math, although it had entered the decade near the bottom of state rankings. Of all states during the 1990s, it was also the most successful in narrowing the minority-white achievement gap (National Education Goals Panel, 1999). In Connecticut, also following steep gains throughout the decade, 4th graders ranked first in the nation by 1998 in reading and math on the NAEP, despite increased poverty and language diversity among its public school students. Its minority-white achievement gap, too, narrowed notably. The proportion of Connecticut 8th graders scoring at or above proficient in reading was first in the nation. In the world, moreover, only top-ranked Singapore could outscore Connecticut students in science (Baron, 1999).

Among the reforms that contributed to such gains were the significant improvements in both states' teaching forces, including in inner cities and rural areas. How did they accomplish this? With ambitious teacher initiatives that introduced standards, incentives and professional learning for teachers, along with curriculum and assessment reforms for schools (Darling-Hammond, 2000a; Wilson, Darling-Hammond, & Berry, 2000).

Notably, neither state succeeded by relaxing teacher education or licensure. On the contrary, they strengthened both. For a teaching license, for example, Connecticut insisted on additional preparation at entry, meaning a major in the content area taught and more pedagogical training as well as learning to teach reading and special-needs pupils and passing basic skills and content tests before entry to teaching. The state also eliminated emergency licensing and toughened requirements for temporary licenses. Teachers must complete a master's degree and a rigorous performance assessment modeled on that of the National Board for Professional Teaching Standards to gain a professional license.

North Carolina likewise increased licensing requirements for teachers and principals (in the form of increased coursework in content and pedagogy as well as licensing tests), required schools of education to undertake professional accreditation through the National Council for Accreditation of Teacher Education (NCATE), invested in improvements in teacher education curriculum, and supported creation of professional development schools connected to schools of education. Both states also developed mentor programs for beginning teachers that extended assistance and assessment into the first year of teaching, and both introduced intensive professional development for veteran teachers. A recent study of North Carolina's reforms noted the strong quality of teachers in the state as a whole and in schools serving diverse student populations. The authors write:

Like the dog that did not bark in the night . . . what is most significant is what is absent. One does not see teachers without pedagogical training, teachers with inadequate content knowledge, or teachers whose own literacy and mathematical skills are poor…. (Asher, et al., forthcoming).

These efforts were successful because both states created strong labor market incentives linked to their teacher standards. Among measures they adopted:

- **Increased and Equalized Salaries, Tied to Standards.** Both states coupled major statewide increases in teacher salaries with improved pay equity across districts. In Connecticut, for example, the average teacher salary climbed from $29,437 in 1986 to $47,823 in 1991, with the equalizing nature of the state aid making it possible for urban districts to compete for qualified teachers. Because Connecticut's state teacher salary assistance could be spent only for fully certified teachers, districts had greater incentives to recruit those who had met the high new standards, and individuals had greater incentives to meet these standards. North Carolina created standards-based incentives by adopting notable salary increases for teachers to pursue National Board Certification, so that North Carolina now has more teachers certified by the National Board than any other state.
- **Recruitment Drives and Incentives.** To attract bright young candidates, both states initiated programs to subsidize teacher education in return for teaching commitments. The highly selective North Carolina Teaching Fellows program, for example, paid all college costs, including an enhanced and fully funded teacher education program, for thousands of high-ability students in return for several years of teaching. After seven years, retention rates for these teachers exceeded 75%, with many of the remaining alumni holding public school leadership posts (NCTAF, 1996). Connecticut's service scholarships and forgivable loans similarly attracted high-quality candidates and provided incentives to teach in high-need schools and shortage fields, while the state also took steps to attract well- trained teachers from elsewhere. By 1990, nearly a third of its newly hired teachers had graduated from colleges rated "very selective" or better in the Barron's Index of College Majors, and 75% had undergraduate grade point averages of "B" or better (Connecticut State Board of Education, 1992, p. 3).
- **Support Systems.** Both states bolstered support systems that make a difference in stemming teacher turnover. North Carolina launched a mentoring program for new teachers that greatly increased their access to early career support (National Education Goals Panel Report, 1998). Connecticut provided trained mentors for all beginning teachers and student teachers as part of its staged licensing process. For existing teachers, North Carolina created professional development academies, a North Carolina Center for the Advancement of Teaching, and teacher development networks such as the National Writing Project and analogous institutes in mathematics. This was in addition to its incentives for National Board Certification. Connecticut, among other things, required continuing professional development, including a master's degree for a professional license.

Such teacher reforms began paying off early on. After Connecticut's $300 million 1986 initiative, for instance, the higher salaries and improved pay equity, combined with the tougher preparation and licensing standards and an end to emergency hiring, swiftly raised teacher quality. An analysis found, in fact, that within three years, the state not only had eliminated teacher shortages, even in cities, but also had created surpluses (Connecticut State Department of Education, 1990). Even as demand increased, the pool of qualified applicants remained solid. A National Education Goals Panel report (Baron, 1999) found that in districts with sharply improved achievement, educators cited the high quality of teachers and administrators as a critical reason for their gains and noted that "when there is a teaching opening in a Connecticut elementary school, there are often several hundred applicants" (p. 28).

These teacher initiatives occurred alongside other education changes—increased investments in early childhood education and in public schools generally, as well as wide-ranging, standards-based reform—which also contributed to the states' student achievement gains. There is little doubt, however, that higher-quality teachers supplied to all schools were substantial contributors to these other reforms as well as to the overall achievement increases. Both states sought to increase not only salaries and the quality of preparation for teachers, but also the incentive structure for distributing teachers to fields and locations. Both sharply reduced hiring of unlicensed and underprepared staff. Most notably, both held to the course of teacher improvement over a sustained period—more than 15 years in each case. They demonstrate what state policy in support of good teaching can accomplish.

B. District Approaches

District success stories reflect the importance of recruiting, inducting and supporting qualified teachers using policy tools available at the local level and leveraging state assistance. Following are just four examples of what urban districts in high-demand states have done.

New York City. New York City illustrates how a focus on recruiting qualified teachers, coupled with necessary salary increases, can have a large effect in a brief period. The city long had hired thousands of underprepared teachers, typically filling as many as half of its vacancies with uncertified applicants, many well after September. The state, however, pressured the city to hire qualified teachers and mandated that uncertified teachers could no longer teach in low-performing schools. This, plus awareness of pending NCLB requirements, led to the improvements. The district focused on more aggressive recruiting and hiring of qualified teachers and implemented a steep increase in salaries—averaging 16% overall and more than 20% for beginning teachers—to make them more competitive with surrounding suburban districts. With these policies, 2002-2003 vacancies were filled by July, and 90% of new hires were certified, up from 60% the year before. The remaining 10% were in programs that would lead to certification by the end of the school year (Hays & Gendar, 2002).

Community School District #2. Much earlier, New York City's Community District #2 was an oasis widely heralded as a turnaround story, with a strategic emphasis on professional development for teachers and principals. But student achievement gains clearly relied on both a development *and* recruitment strategy (Elmore & Burney, 1999). In 1996, after a decade of reforms focused on strengthening teaching, this "majority minority" district—which serves large numbers of low-income and immigrant students—realized sharp achievement gains that ranked it 2nd in the city in reading and math.

Sweeping changes instituted by Superintendent Anthony Alvarado stressed continuing professional development for teachers and principals, coupled with a relentless concentration on instructional improvement. At the same time, Alvarado recognized the need for more talented and committed teachers and principals. Backed by the teachers' union, he replaced nearly half the teacher workforce and two-thirds of principals over a period of years through a combination of retirements, pressure and inducements. Meanwhile, the central office carefully managed the recruitment, hiring and placement of new teachers and principals. It ended the hiring of unprepared teachers and sought recruits from several leading teacher education programs in the city, forging partnerships for student teaching and professional development with these institutions as well. Similar programs for developing principals were launched. The district's growing reputation for quality also attracted other teachers. Salary changes were not within the district's purview. Its strategies, rather, involved recruiting aggressively, creating university partnerships to develop a pipeline of well-prepared teachers, and supporting teachers with strong mentoring and professional development.

New Haven, California. California success stories are particularly notable because that state in recent years has ranked first in the nation in the number of unqualified teachers. In this high-demand context, with state policies that were, until recently, relatively unsupportive (e.g., low expenditures, lack of reciprocity with other states, restricted teacher education options), some districts have nonetheless achieved significant staffing improvements. New Haven Unified School District, just south of Oakland in Union City, which enrolls 14,000 mostly low-income and minority students, is one that has succeeded while neighboring districts have not. New Haven combined high salaries, aggressive recruiting and close mentoring with a high-quality training program worked out with area universities. Although not a top-spending district, it invested its resources in teacher salaries and good teaching conditions. In 1998, for example, New Haven's salaries were more than 30% higher than nearby Oakland's, where large numbers of unqualified teachers were hired, even though New Haven's per-pupil spending was below Oakland's (Snyder, 2002).

Thus, over an extended period it built a well-prepared, highly committed and diverse teaching staff. For the 2001-2002 school year, 10 of its 11 schools had no uncredentialed teachers. The district averaged 0.1% uncredentialed teachers—while some neighboring districts averaged more than 20% (Futernick, 2001). New Haven uses advanced technology and a wide range of teacher supports to recruit from a national pool of exceptional teachers and to hire them quickly. The district was one of California's first to implement a Beginning Teacher Support and Assessment Program that assists teachers in their first two years in the classroom; all beginning teachers get help from a trained mentor, who is given release time for the purpose. In addition, New Haven collaborated with California State University- Hayward on the right kind of alternative-certification program, combining college coursework and an internship, including student teaching, conducted under the close supervision of university- and school-based educators. As a result of these initiatives, the district has a teacher surplus in the midst of general shortages (for details, see Snyder, 2002).

San Diego, California. Using similar strategies, San Diego City Schools recently overhauled its teacher recruitment and retention system, aggressively recruiting well-trained teachers, collaborating with universities on new training programs in high-need fields, and creating smooth pathways with local schools of education. It offers contracts to well-prepared teachers as early as possible (sometimes as much as a year in advance of hiring) and reaches out to teachers in other states. In addition, the district streamlined the hiring process, putting the entire system online, improving its capacity to manage hiring data, vacancy postings and interviews that had slowed the process and caused many candidates to give up and go elsewhere. In the fall of 2001, districts like San Francisco and Los Angeles hired hundreds of uncredentialed teachers, and the state as a whole hired more than 50% of novices without full credentials. But San Diego filled almost all of its 1,081 vacancies with credentialed teachers, eliminating all but 11 of the hundreds of previously hired emergency permit teachers who had been assigned largely to high-minority, low- income schools. (Darling-Hammond, et al., 2002).

What State and Local Successes Tell Us

Taken together, these state and local cases demonstrate that determined, well-focused, and sustained efforts can make a difference in staffing even hard-to-staff schools, which in turn greatly increases the probability of student learning. These cases also make clear that schools can be staffed without lowering the bar on teacher standards by counting untrained novices as "highly qualified" or by encouraging states to dilute certification requirements. While it is important to broaden the sources of supply for teaching, it is also essential to safeguard the quality of that supply if the NCLB goals for children's learning are to be achieved. This can be achieved by clarifying three aspects of the law:

- Teachers should be considered "fully certified" under NCLB's definition of "highly qualified" when they have *completed* a traditional or alternative-route program.
- "Full certification" should continue to include content and pedagogical preparation.
- Standards should be adopted for acceptable alternate-route (and traditional) programs. One careful synthesis of teacher preparation research (Wilson, Floden, & Ferrini-Mundy, 2001, p. 30) suggests, for example, that the following components should be included in high- quality, alternate-certification programs (components that could be applied equally as well to traditional programs):
    - High entrance standards.
    - Intensive training in instruction, management, curriculum, assessment and how to work with diverse students.
    - Extensive mentoring and supervision by well-prepared teachers.
    - Frequent and substantial evaluation.
    - Guided practice in lesson planning and teaching, with benchmarks for competence prior to taking full responsibility as a teacher.
    - High exit standards tied to state standards for teaching.

Around such standards states and districts can improve teacher preparation, with Washington developing incentives to attend such programs, thereby boosting supply while encouraging the elimination of ineffective alternatives.

-

VI. The Need for a National Teacher Supply Policy

While we can learn a good deal from state and local successes, such cases are the exceptions to the rule. They stand out amid widespread use of underprepared teachers and untrained aides, mainly for disadvantaged children in schools that suffer from poor working conditions, inadequate pay and high teacher turnover. Thus while much that must be done lies at the state, district and even school level, the federal government has a critical role to play, focused on three goals:

- Enhancing the supply of qualified teachers targeted to high-need fields and locations.
- Improving retention of qualified teachers, especially in hard-to- staff schools.
- Creating a national labor market by removing interstate barriers to mobility.

This can be accomplished, we believe, by drawing in part on the federal experience with medical manpower programs. Since 1944, Washington has subsidized medical training and facilities to meet the needs of underserved populations, to fill shortages in particular fields and to increase diversity in the medical profession. (Note 20) The federal government also collects data to monitor and plan for medical manpower needs. This consistent commitment, on which we spend hundreds of millions of dollars annually, has contributed significantly to America's world-renowned system of medical training and care. Although the teacher labor market is also vital to the nation's future, federal efforts in this area have tended to be modest, fragmented and inconsistent over time. (Note 21) Washington has periodically adopted programs to enhance teacher supply, but these have not continued on the scale and with the targeting needed to address the problems noted. There has been little investment in developing a national system to monitor and adjust the teacher labor market. (Note 22) There have been scarce efforts to develop the capacity of training institutions to ensure teacher supplies in high-demand locales and fields. There has been no serious attempt to establish comprehensive federal-state partnerships like those created to meet specific health-field shortages. Thus we recommend a series of measures to create a federal teacher supply program that substantially addresses the problems we face. The general strategy is to supply grants to individuals and institutions, with funds concentrated where they are needed most, where they will create new institutional capacity, and where they will work to relieve the maldistribution of teaching talent.

Increasing Supply in Shortage Fields and Areas

While there have long been surpluses of candidates in elementary education, English, and social studies in most states, there are inadequate numbers of teachers trained in high-need areas like mathematics, physical science, special education, bilingual education and English as a Second Language (ESL). The nation requires targeted incentives to attract qualified teachers to schools and areas that historically have been undersupplied. A two-pronged approach seems warranted. First, Washington should consolidate all of its small-scale fellowship, scholarship, and loan forgiveness programs into a single, sustained program of service scholarships and forgivable loans that includes the following elements:

- Scholarships allocated both on the basis of academic merit and indicators of potential success in teaching, such as perseverance, capacity and commitment.
- Scholarships targeted in substantial part to areas of teaching shortage. Washington would allocate half the funds to national priorities, reserving the other half for states to establish their own priorities.
- Scholarships awarded in exchange for teaching in priority schools, defined on the basis of such criteria as poverty rates and the percentage of language minority students.
- Awards available for training at either the undergraduate or graduate level, with scholarships forgiven over three to five years in exchange for teaching in high-need areas and fields. Because the chances of staying in teaching increase significantly after three years, calibrating the length of the service required with an inducement of sufficient size would be important to the initiative's success.

The federal government is the appropriate primary source of such programs for two reasons. First, the program can influence the flows of talent across areas of the country. Second, the budgetary implications are extremely modest for Washington relative to the states. This is an area where a relatively small federal outlay can go a long way—and actually save the nation sizable sums.

Assume, for example, that the country needs an annual influx of 40,000 new teachers supported by such scholarships (Note 24) and that each candidate would receive up to $20,000 to cover tuition for undergraduate or graduate teacher preparation. (Note 25) Such a program, costing $800 million a year, could meet most of the nation's teacher supply needs in a few years. Given that we currently lose billions of dollars each year due to early attrition from teaching—much of it a result of hiring underprepared teachers—this program would repay itself many times over if it induced recipients to remain in teaching even for several more years.

Such a program alone, however, would not be sufficient to attack the systemic nature of teacher shortages in urban and isolated rural schools. Recall that teacher labor markets are intensely local and that many young teachers have a strong preference to teach close to home, hurting some districts' efforts to attract qualified applicants. Urban and rural schools must either lure applicants from other areas, which is often difficult, or enhance the pool of college graduates who grew up in neighborhoods served by urban schools. This second prospect suggests a recruitment strategy that might underwrite the development of "grow your own" programs in urban and rural areas. (Note 26)

Grants are needed to build the capacity of teacher preparation programs within cities where the problems are most severe. These programs would need to meet three criteria: ensuring a high-quality teacher preparation experience, attracting local residents to the programs, and ensuring a pipeline from preparation to hiring.

Some cities have many higher education opportunities, but not all are affordable to local residents or have close ties with the district to facilitate an easy pathway from preparation to hiring. The value of many alternate-route programs is that they finance and prepare candidates explicitly for a given district; thus the district reaps the investment's benefits, and candidates know they will have a job. When these are high-quality programs with the components described earlier, the bargain is a good one. Some cities, like New York and San Diego, have created local university partnerships that include underwriting the preparation of candidates, with service in the city's schools required in exchange. Some of these universities enable candidates to engage in practice teaching in professional development schools that are particularly successful with urban and minority teachers, so that they learn effective practices rather than mere survival. And some programs target both local residents and longtime paraprofessionals already knowledgeable about and committed to their communities. The key is a combination of strong training targeted at local talent and strong incentives for hiring and retention in the district.

Such opportunities could be encouraged by a new federal grant program, possibly with a state or local matching requirement, directed to urban universities and districts to create or expand programs that meet the standards for program quality that we have described and that support local candidates from preparation through hiring. Some funds could be used for program development or expansion, while others could provide subsidies to enable candidates to attend, with pledges for service in the district. Analogs are available in federal support for urban medical training models (see Appendix A). [from ldh: Note I use federal instead of Washington because Washington is a city or a state, NOT a level of government].

If we wanted these institution-building grants to operate in the 50 largest cities, with an average of two programs per city (calibrated to size and need), and if each developmental grant allocated $1 million per program for each of five years, the annual cost would be $100 million (with attendant administrative and evaluation costs adding marginally to this sum). This would add only modestly to the previously noted scholarship program and still keep total yearly expenses far below the noted savings. If we wanted to spread these costs over time, moreover, the programs could be phased in over, say, a decade.

The models that emerged might well be richly diverse, including new forms of professional development schools that emulate the teaching hospitals used to develop state-of-the-art medical practices. They might include new applications of distance technology, new forms of collaboration by the private and public sectors, and new kinds of partnerships among schools, districts and the multiple universities. This would make the investment worth its weight in gold, especially given the subsequent diffusion of successful models.

Improving Teacher Retention

In addition to incentives for entering teaching, improving teacher retention in high-need areas would be an essential goal of a federal teacher supply program for education. Growing evidence indicates that high turnover, particularly in the early years, is a major part of the problem for the system, especially in hard-to-staff schools. Washington could help stem such attrition by becoming engaged in several areas, starting with helping to ensure that teachers in such schools receive appropriate preparation and mentoring.

a. Preparation and Support. While quality local programs to prepare urban teachers would go a long way toward supplying schools, a great unfinished task in American education is to create conditions for better support of new teachers, encompassing hiring procedures, protected initial assignments, steady provision of mentor and other support, and improved evaluation to help novices. These matters have been neglected for too long, and they particularly harm hard-to-staff schools that need greater personnel stability if they are to create effective learning communities. The intervention point here clearly is induction, beginning with hiring and assignment practices, reduced teaching loads, close fits between qualifications and teaching duties, and the orchestration of support from experienced teachers and administrators. How might more effective induction practices be promoted?

State certification policy is one vehicle. As evidenced by such cases as Connecticut, states can establish conditions for effective induction through certification requirements established for new teachers. In addition to encouraging such innovations through the U.S. Department of Education's leadership activities, Washington could create a targeted, matching grant program aimed explicitly at supporting effective induction practices. Since many states and some districts have created induction programs, some resources already are focused on these needs. Relatively few programs, however, ensure that expert mentors in the same teaching field are made available for in- classroom support, the component of induction with the greatest effect on teacher retention and learning.

Part of such a program would supply grants to state agencies willing to develop statewide induction programs that would be integrated with their licensure and certification requirements. States might use such grants to fund universities, districts and other agencies to develop and test model induction programs, concentrating on support for new teachers in hard-to-staff schools. Another part of the program would distribute grants to high-need districts to support induction practices such as mentor cadres and related supports.

The annual costs would again be exceedingly modest. The grants to states might supply startup funds, with the pledge that states would continue effective programs and practices after that period. If individual state grants averaged $500,000 annually for three years running and were phased in 10 states at a time, the total direct cost of this part would be $75 million, allocated over seven years. The grants to local districts might allocate an average of $250,000 a year for three years of startup funds, also with the requirement that districts continue effective practices. If 100 district grants were given to 20 districts a year and phased-in over time, the second part would total $75 million, also spread over seven years. If Washington took on the role of evaluating and disseminating knowledge from these programs, the nation would benefit considerably from new policies and practices that receive hardy tests under a variety of conditions.

b. Pay and Working Conditions. These factors clearly are of great importance, as is evident from states and localities that have implemented successful policies directed at salaries, benefits and working conditions. Too many urban districts are doubly disadvantaged in the competition for teaching talent. They have difficult living and working conditions, and they offer salaries below those of nearby suburban districts. As noted, Connecticut provides an example of how a state dealt with these problems by both raising and equalizing salaries.

While issues of pay and working conditions are centrally the concerns of states and localities, Washington could encourage more states to address these issues by sponsoring research within and across states on the success of various strategies in different contexts. These might include systemic state strategies like Connecticut's and local experiments with compensation plans. Experiments with extra pay for teaching in hard-to-staff schools (sometimes known as "combat pay") generally have proven ineffective, but some states and districts are exploring further innovation with compensation and working conditions that bear watching. For example, some analysts have advocated advancement on teacher salary schedules based on indicators of performance in teaching, including National Board Certification and other measures of merit or accomplishment. California implemented $10,000 bonuses for National Board-certified teachers, increased to $30,000 for such teachers who taught in low-performing schools. California also implemented its Teachers as a Priority Program, which sent resources to high-need schools to recruit and retain fully certified teachers through improving working conditions, adding mentors, reducing class sizes and providing hiring bonuses. Moreover, hard-to-staff districts might experiment with pay packages that include, for example, special housing, parking, or transportation allowances, additional medical and retirement benefits, or summer-based professional development opportunities for travel, workshops, institutes and other experiences.

In addition to sponsoring research, Washington might play a role in stimulating the development and testing of innovative compensation and support models explicitly designed to retain effective teachers in needy schools and districts. In this case, the Department of Education or other relevant agency would announce a national grant program that would support two phases of work, the first to develop innovative compensation plans, the second to evaluate trials of these models to determine their effectiveness. If 10 to 12 such grants were let, then studied over a significant period, the knowledge return could be substantial, leading to the adoption of new compensation practices in districts that historically have had difficulty retaining teachers. Once evaluation research had validated the worth of such models, there would be a basis for states and districts to invest in these models out of operating funds.

c. The Prospect of Success. Finally, teachers are more likely to stay in schools where they feel they can succeed. In this regard, research stresses the importance of professional supports and redesigned schools to build stronger teacher-student relationship that promote trust, motivation, commitment and collective efficacy (for one example, see Bryk & Schneider, 2002). These "soft" features of schools are alterable through more skillful management and organization, which could be supported through development of new administrative leadership programs and continued support of redesigned schools, such as those offered through the New American Schools development program and the Small Schools Act.

Teachers in difficult classrooms, however, are unlikely to be encouraged to stay by the perverse incentives that may be encouraged by NCLB. Under that law, schools are being branded as "in need of improvement"—widely viewed as a euphemism for "failing"—if all students and such subgroups as poor, minority and limited-English-proficient students do not all show adequate yearly progress on test scores. Schools stand to be reconstituted and states and districts stand to lose funds based on missing testing targets. The problem is not only that school scores are so volatile as to be useless as indicators of improvement (Linn & Haug, 2002) and that the targets adopted are likely to result in more than half the nation's schools seen as failing over the next few years. The problem is also that the stigma is likely to make it even harder for such schools to recruit and retain highly qualified teachers. These labels and the accompanying pressure could chase teachers away from such schools even more persuasively than current conditions (Clotfelter et al., 2003; Figlio, 2001; DeVise, 1999; Tye and O'Brien, 2002).

If evidence mounts that schools face a teacher exodus because they are seen as failing or because of rising dismay at excessive accountability pressures, countervailing measures may be necessary. In addition to amending NCLB to develop more sensible measures of progress, Washington, along with states and localities, may need to create other inducements to teach, and remain teaching, in such schools. Otherwise, even less-qualified individuals may end up instructing these students.

3. Facilitating a National Labor Market for Teachers

Finally, Washington can create the foundation of a national labor market for teachers, including the removal of unnecessary interstate barriers to teacher mobility. Because teacher supply and demand vary regionally, the country can only benefit if states with teacher surpluses in particular fields can be connected to states with corresponding shortages. Washington could work with states to accomplish three goals:

- Developing common licensing exams and interstate agreements about content and pedagogical coursework that would facilitate reciprocity and respond to the standard called for by NCLB.
- Creating a system of pension portability across the states.
- Providing labor market data and analyses for federal, state and local planning.

Several groups already are working on these agendas in ways that could be leveraged toward genuine changes. For example, the Interstate New Teacher Assessment and Support Consortium, sponsored by the Council of Chief State School Officers, has brought together more than 30 states to create licensing standards and new assessments for beginning teachers. The consensus they have forged could be the basis for an eventual national system. The organization of State Higher Education Executive Officers, along with the Education Commission of the States, has examined how to achieve teacher pension portability, and TIAA-CREF has developed such plans as well. A public/private partnership to stimulate the next steps in these plans could be extremely productive.

Finally, the long-standing federal role of keeping statistics and managing research is well suited to the job of creating a database and analytic agenda for monitoring teacher supply and demand. Such a system, which would inform all other policies, could document and project shortage areas and fields, determine priorities for federal, state and local recruitment incentives, and support plans for institutional investments where they are needed.

In making all of these recommendations, we are mindful of the federal deficits that are looming. However, these initiatives could be undertaken for less than 1% of the $350 billion tax cut enacted in May 2003, and, in a matter of only a few years, they would build a strong teaching force that could last decades. We would stress again, moreover, that these proposals could save far more than they would cost. The savings would include the several billion dollars now wasted because of high teacher turnover as well as the costs of grade retention, summer schools and remedial programs required because too many children are poorly taught. This is to say nothing, moreover, of the broken lives and broader societal burdens that could be avoided with strong teachers in the schools that most need them. In the competition for educational investment, the evidence strongly points to the centrality of teacher quality to educational improvement. That should be a centerpiece of the nation's education agenda. The benefits of this strategy, in terms of students' school success, employability and contributions to society, will, we believe, repay the costs many times over.

Notes

0. The research reported here was originally commissioned by the Education Commission of the States as part of its 10th Amendment Project. A version of the report can be found at http://www.ecs.org/clearinghouse/46/34/4634.doc. The authors wish to acknowledge the assistance in preparing this report provided by Debbi Harris and Lisa Ray of Michigan State University, and Lisa Marie Carlson of Stanford University. The opinions expressed, however, are the authors alone.

- 1. A number of recent polls demonstrate that large majorities of parents and members of the general public (90%) believe that getting and supporting well-qualified teachers is the strategy most likely to improve schools; that such teachers should have knowledge of content, how children learn, and how to teach; and that salaries – and taxes – should be raised if necessary to get well-qualified teachers in all schools. See, for example, Educational Testing Service, 2002; Recruiting New Teachers, 1998.
- 2. In education, including in the NCLB legislation, "licensure" and "certification" often are used interchangeably. However, in most other professional fields, licensure refers to state requirements governing entry to a field, while certification denotes advanced standing based on standards set by a profession. For example, states grant physicians a license to practice medicine; professional boards grant certification in particular medical specialties. Similarly, the National Board for Professional Teaching Standards (NBPTS) certifies teachers who demonstrate "accomplished practice," while states grant licenses to practice. Here, we conform to general usage, using "certification" and "licensure" as equivalent terms for the mandatory state requirements for entry to teaching.
- 3. See Darling-Hammond and Youngs (2002) for a review of the evidence on which the report's recommendations are based.
- 4. 34 CFR Part 200, Federal Register, Vol. 67, No. 231 (December 2, 2002), p. 71712. Downloaded on April 20, 2003 from http://www.ed.gov/legislation/FedRegister/finrule/2002-4/120202a.html.
- 5. 34 CFR Section 200.56, Federal Register, Vol. 67, No. 231 (December 2, 2002), p. 71765.
- 6. 34 CFR Section 200.56, Federal Register, Vol. 67, No. 231 (December 2, 2002), p. 71764.
- 7. HB 318. Downloaded on 5/3/03 from http://www.capitol.state.tx.us/tlo/78R/billtext/HB00381.thm.
- 8. See *Teacher Quality, Introduction,* Thomas B. Fordham Foundation Website, http://www.edexcellence.net/topics/teachers.html.
- 9. Some opponents of teacher certification have misconstrued one finding of this study to argue against teacher education requirements. Because students of a small number of science teachers with temporary or emergency certification (24 of 3,469 teachers in the sample) did no worse than students of certified teachers, these opponents have termed teacher certification unnecessary (see e.g. Strayhorn, 2003). However, these teachers, like those with standard certification, were found to be more effective than uncertified teachers. Another analysis of these data (Darling-Hammond, Berry, & Thoreson, 2001) showed that most of the science teachers in the sample with temporary or emergency certificates had many years of experience and subject matter and education training comparable to that of certified teachers. Their backgrounds and teaching contexts suggested that many were previously certified, out-of-state entrants working on a temporary credential while becoming certified in a new state. Others were certified in math or another sub- field of science. It is not surprising, then, that their students did about as well as those of certified teachers with similar qualifications. Only a third of the NELS sample teachers on temporary/emergency licenses were new entrants to teaching with little education training. The students of this sub-sample had smaller achievement gains than those of the more experienced, traditionally trained teachers in an analysis of covariance that controlled for pre- test scores, content degrees and experience.
- 10. A decade ago, only a quarter to a third of newly hired teachers were "newly minted." This proportion has increased with growing demand, reaching as many as half of new hires in the late 1990s. In a few high- demand states like California, the proportion has reached 60 percent, but this is unusual.
- 11. Various studies of teacher supply have found that about 20% to 30% of teachers who have left the classroom eventually return to teaching in the same state (Beaudin, 1993; Massachusetts Institute for Social and Economic Research, 1987; Murnane et al., 1991). Some leave to teach in another state, although most studies have not had data sets to track these individuals. The likelihood that those who have left teaching will re- enter depends heavily on salary levels and work conditions (Beaudin, 1993; 1995).

- 12. Boe (1998) and colleagues have found that, nationally, delayed entrants comprise about a third of new entrants to teaching annually, which in this case would be about 50,000.
- 13. Between 1983 and 1999, annual graduates with a bachelor's or master's degree in education jumped from 134,870 to 234,408. However, this does not translate directly into new teacher supply, since bachelor's degrees in education now represent fewer than half of newly prepared teachers. Most now receive a degree in a disciplinary field and a second major, minor or master's in education. While a growing share of teachers are trained in master's programs, many master's degrees are gained after teachers have already completed initial preparation.
- 14. Because a large majority of alternative programs are run by or in collaboration with universities, their graduates are counted in university totals. Estimates of alternative-certification programs vary, depending on classification, but by 1999, 40 states and the District of Columbia had 117 state-authorized programs (Feistritzer & Chester, 2000). In addition, the American Association of Colleges for Teacher Education (1996) cataloged 328 alternative programs run by colleges and universities.
- 15. Researchers using longitudinal data from the 1993-94 Baccalaureate and Beyond survey find a 4-year attrition rate of about 20% for those who entered teaching directly after college (Henke, et al., 2000). Ingersoll (2001) extrapolates from cross-sectional data on teacher attrition (from the 1999-2000 Schools and Staffing Surveys) to estimate a 5-year attrition rate for beginning teachers of 46%, including private school teachers. However, the 5-year attrition rate for public school teachers is only about 38%. Furthermore, some individuals who left teaching for childrearing or further study will have returned to the classroom in the first five years—a proportion that, other estimates suggest, could be about 20% of leavers. With this adjustment, the 5-year cumulative attrition rate would be just over 30% for public school teachers.
- 16. Analysts have long recognized that salary differentials across teaching areas contribute to shortages, based on the sensible proposition that individuals are influenced by salaries available to them. In response, some have argued for altering the structure of public school salary schedules by allowing differential pay across teaching specialties. Some experiments along these lines have appeared over the years, including recent efforts in Cincinnati, OH, Douglas County, CO, and Denver, CO, among others. In 2003, the Kentucky State Department of Education awarded grants to ten districts for innovations in salary systems. These experiments are worth careful study, but for the most part salary schedules have remained uniform and fixed. See Kershaw & McKean (1962) and Murnane, et al. (1991) for further discussion of this issue.
- 17. Flanagan and Grissmer (2002) point out, however, that while one-third of the inequality in educational spending is within-state, almost two-thirds of the variance is between-state. Even accounting for between-state differences in the costs of education, this basic fact points to the need for equity policies at the federal level.
- 18. For example, a study for the California Commission on Teacher Credentialing (CCTC, 1998) documented the difficulties out-of-state candidates experienced in seeking teaching positions. Problems included costs of courses and exams, confusion about how to complete the many and varied requirements, and redundancy with other requirements teachers had already met elsewhere. In a survey of out-of-state teachers who had received an initial permit to teach in California, credential requirements were the leading factor in decisions to leave the state.
- 19. In the fall of 1999, Connecticut had 30% students of color, including the 12th largest Hispanic enrollment in the nation, and in 2002, 36% of students attended Title I schools. In the same years, North Carolina had 38% students of color, including the 8th largest enrollment of African Americans, and 38% of students attended Title I schools (NCES, 2001, table 42; NAEP State Data, 2002, retrieved from http://nces.ed.gov/nationsreportcard/statedata).
- 20. See Appendix A for a brief history of federal involvement in medical teacher supply policy.
- 21. See Appendix B for a brief history of federal involvement in the teacher labor market.
- 22. Although the Schools and Staffing Surveys provide useful data for monitoring aspects of supply and demand, they have never been fully exploited for this purpose. Modifications to the questionnaires have made the data about training and certification too imprecise for some important analyses. Furthermore, the delay between surveys and the delay in releasing data to the public for outside analysis make them much less useful than they could be for monitoring supply trends. Although the SASS was intended to occur every three years, the delayed 1999-2000 survey came six years after the 1993-94 survey.
- 23. While Title I status is a key indicator, the Title I program fails to reach a large portion of students from poor families. Thus a national program of teacher scholarships ultimately should be tied to service targeted at the actual distribution of poor children, not to Title I school status alone.
- 24. Of the 250,000 teachers hired annually, no more than 50,000 enter without standard certification in their main teaching field. This overestimates the need, since many of these teachers are certified in some field, if not the one they are teaching, and some are in transition from one state to another or have been hired without yet taking the state licensing examinations, so they are only temporarily in the 'not fully qualified' category.
- 25. This is enough to pay full tuition and some stipend in a public college or university for a one-year master's program in teaching for recent graduates or mid-career entrants—or enough for 2-3 full years of undergraduate tuition in a state university for juniors and seniors preparing to teach.
- 26. We are indebted to Susanna Loeb for suggesting this point, and for elaborating it in several papers she has written with her colleagues.
- 27. This might include school improvement measures that rely on aggregated longitudinal scores for individual students, rather than annual cross-sectional estimates that can fluctuate from year to year for a variety of reasons unrelated to school practices; averages of these longitudinal score gains over multiple years; annual targets that are not statistically unreasonable; and multiple measures of school practice and performance that extend beyond test scores.

References

American Association of Colleges for Teacher Education. (1996). *Alternative paths to teaching: A directory of post-baccalaureate programs.* Washington, DC: Author.

American Association for Employment in Education. (2000). *Educator supply and demand in the United States: 2000 Research Report.* Columbus, OH: Author.

Andrew, M. & Schwab, R. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of eleven teacher education programs. *Action in Teacher Education, 17,* 43-53.

Asher, C., (forthcoming). Improving teaching in North Carolina: How school design intersects with school reform. Seattle: Center for the Study of Teaching and Policy, University of Washington.

Baron, J. (1999). *Exploring high and improving reading achievement in Connecticut.* Washington, DC: National Educational Goals Panel.

Ballou, D. (1996). Do public schools hire the best applicants? *Quarterly Journal of Economics 111*(1), 97-133.

Ballou, D., & Podgursky, M. (1997). *Teacher pay and teacher quality.* Kalamazoo, MI: The Upjohn Institute for Employment Research.

Baugh, W.H., & Stone, J.A. (1982). Mobility and wage equilibration in the educator labor market. Economics of Education Review, 2 (3): 253- 274.

Beaudin, B. (1993). Teachers who interrupt their career: Characteristics of those who return to the classroom. *Educational Evaluation and Policy Analysis, 15,* (1): 51-64.

Beaudin, B. (1995). Former teachers who return to public schools: District and teacher characteristics of teachers who return to the districts they left. *Educational Evaluation and Policy Analysis, 17*(4): 462-475.

Begle, E. (1979). *Critical variables in mathematics education: Findings from a survey of empirical literature.* Washington, DC: National Council for Teachers of Mathematics.

Benner, A. D. (2000). *The cost of teacher turnover.* Austin, TX: Texas Center for Educational Research.

Betts, J. R., Rueben, K.S., & Danenberg, A. (2000). *Equal resources, equal outcomes? The distribution of school resources and student achievement in California.* San Francisco: Public Policy Institute of California.

Bliss, T. (1992). Alternative certification in Connecticut: Reshaping the profession. Peabody Journal of Education, 67(3): 35-54.

Boe, E., Bobbitt, S.A., Cook, L.H., Barkanic, G., Maislin, G. (1998). *National trends in teacher supply and turnover for special and general education.* Data analysis report no. 1998- DAR1. Philadelphia: Center for Research and Evaluation in Social Policy, University of Pennsylvania.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2003, March). *The draw of home: How teacher preferences for proximity disadvantage urban schools.* Unpublished paper.

Brewer, D. (1996). Career paths and quit decisions: Evidence from teaching. *Journal of Labor Economics, 14*(2), 313-339.

Bridges, E. (1990). Evaluation for tenure and dismissal. In J. Millman & L. Darling-Hammond (Eds.). *Handbook of teacher evaluation* (pp. 147-157). Newbury Park, CA: Sage Publications.

Bryk, T. & Schneider, B. (2002). *Trust in schools: A core resource for improvement.* New York: Russell Sage Foundation.

California Commission for Teacher Credentialing (CCTC) (1998a, June). A report on issues involving the recruitment and retention of teachers prepared in other states. Sacramento, CA: Author.

California State University. (2002a). *First system wide evaluation of teacher education programs in the California State University: Summary Report.* Longbeach, CA: Office of the Chancellor, California State University.

California State University. (2002b). *Preparing teachers for reading instruction (K-12): An evaluation brief by the California State University.* Longbeach, CA: Office of the Chancellor, California State University.

California Select Committee on Low Performing Schools. (2001).

Carnegie Forum on Education and the Economy. (1986). *A nation prepared: Teachers for the twenty-first century.* New York: Carnegie Forum, Carnegie Corporation of New York.

Carroll, S., Reichardt, R. & Guarino, C. (2000). *The distribution of teachers among California's school districts and schools.* Santa Monica, CA: RAND Corporation.

Choy, S., Henke, R., Alt, M., Medrich, E., & Bobbitt, S. (1993). *Schools and staffing in the United States: A statistical profile, 1990-91.* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Clotfelter, C., Ladd, H., Vigdor, J., & Diaz, R. (2003, February)(revised). *Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teachers?* Paper prepared for the annual meeting of the American Economic Association, Washington, DC.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, R., et al. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.

Connecticut State Department of Education (1990). *Impact of Education Enhancement Act, Research Bulletin, School Year 1990, No. 1.* Hartford, CT: Author.

Connecticut State Board of Education. (1992). *The other side of the equation: Impact of the teacher standards provisions of the Education Enhancement Act.* Hartford, CT: Author.

Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching.* New York: National Commission on Teaching & America's Future (0-9654535-3-7).

Darling-Hammond, L. (2000a). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1): http://epaa.asu.edu/epaa/v8n1.

Darling-Hammond, L. (Ed.) (2000b). *Studies of excellence in teacher education (three volumes).* Washington, DC: American Association of Colleges for Teacher Education.

Darling-Hammond, L. (in press). Access to quality teaching: An analysis of inequality in California's public schools. *Santa Clara Law Review.*

Darling-Hammond, Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis, 23*(1), 57-77.

Darling-Hammond, L., Chung, R, & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education, 53*(4): 286-302.

Darling-Hammond, L., Hightower, A., Husbands, J., LaFors, J., & Young, V. (2002, April). *Building instructional quality: Inside-out, bottom-up, and top-down perspectives on San Diego's school reform.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Darling-Hammond, L., Kirby, S., & Hudson, (1989). *Redesigning Teacher Education: Opening the Door for New Recruits to Science and Mathematics Teaching.* Santa Monica, CA: RAND Corporation, 1989.

Darling-Hammond, L. & Youngs, P. (2002). Defining "Highly Qualified Teachers:" What does "Scientifically-Based Research" Actually Tell Us? *Educational Researcher, 31*(9), 13-25.

DeVise, D. (1999). A+ Plan Prompts Teacher Exodus in Broward County. *Miami Herald*, Nov. 5, 1999.

Educational Testing Service. (1991). *The state of inequality.* Princeton, NJ: Author.

Educational Testing Service (2002). *A national priority: Americans speak on teacher quality.* Princeton, NJ: ETS.

The Education Trust (2002). *The funding gap: Low income and minority students receive fewer dollars.* Washington, DC: Author. http://www.edtrust.org/main/documents/investment.pdf.

Elmore, R. & Burney, D. (1999). Investing in teacher learning: Staff development and instructional improvement. In L. Darling-Hammond and G. Sykes (Eds.). *Teaching as the learning profession. Handbook of policy and practice* (pp. 263-291). San Francisco: Jossey-Bass.

Feistritzer, E. & Chester, D. (2000). *Alternative teacher certification: A state-by-state analysis 2000.* Washington, DC: National Center for Education Information.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal of Legislation, 28*(2), 465-498.

Ferguson, R.F. & Ladd, H.F. (1996). How and why money matters: An analysis of Alabama schools. In H. Ladd (Ed.). *Holding schools accountable* (pp. 265-298). Washington, DC: Brookings Press.

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives, 7*(9): http://epaa.asu.edu/epaa/v7n9.html

Figlio, D. (2001). *What might school accountability do?* Cambridge, MA: National Bureau of Economic Research.

Flanagan, A. & Grissmer, D. (2002). The role of federal resources in closing the achievement gap. In T. Loveless and J. Chubb (Eds.).

Bridging the achievement gap (pp. 199-226). Washington, DC: Brookings Press.

Fowler, C. (2002). Fast track…slow going? Research Brief. *Ed Policy 2*(1). Retrieved September, 2002 from http://www.edpolicy.org/publications/documents/updatev2i1.pdf.

Fuller, E. (1998). *Do properly certified teachers matter? A comparison of elementary school performance on the TAAS in 1997 between schools with high and low percentages of properly certified regular education teachers.* Austin: The Charles A. Dana Center, University of Texas at Austin.

Fuller, E. (2000). *Do properly certified teachers matter? Properly certified algebra teachers and Algebra I achievement in Texas.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Futernick, K. (2001, April). *A district-by-district analysis of the distribution of teachers in California and an overview of the Teacher Qualification Index (TQI).* Sacramento: California State University, Sacramento.

Gomez, D., & Grobe, R. (1990). *Three years of alternative certification in Dallas: Where are we?* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Goe, L. (2002, April). Legislating equity: The distribution of emergency permit teachers in California. *Educational Policy Analysis Archives, 10*(42). Retrieved November 8, 2002 from http://epaa.asu.edu/epaa/v10n42.

Goldhaber, D.D. & Brewer, D.J. (2000). Does teacher certification matter? High school certification status and student achievement. *Educational Evaluation and Policy Analysis, 22,* 129-145.

Good, T.L. & Brophy, J.E. (1995). *Contemporary educational psychology, 5th edition.* NY: Longman, Inc.

Gormley, M. (2003, January 2). State: The teacher shortage is abating. *Boston Globe.*

Greenwald, R., Hedges, L., & Laine, R.D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66,* 361-396.

Grey, L., Cahahlan, M., Hein, S., Litman, C., Severynse, J., Warren S., Wisan, G., & Stowe, P. (1993). *New teachers in the job market. 1991 update.* Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement.

Gritz, M., & Theobold, N. (1996). The effects of school district spending priorities on length of stay in teaching. *Journal of Human Resources, 31,* 477-512.

Hanushek, E. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84-117.

Hanushek, E. (1996). *School resources and achievement in Maryland.* Baltimore: Maryland State Department of Education.

Hanushek, E., Kain, J., & Rivkin, S. (1998). *Teachers, schools, and academic achievement.* (Working Paper 6691). Cambridge, MA: National Bureau of Economic Research.

Hanushek, E., Kain, J., & Rivkin, S. (1999). *Do higher salaries buy better teachers?* Cambridge, MA: National Bureau of Economic Research.

Hanushek, E., Kain, J., & Rivkin, S. (2001). *Why public schools lose teachers* (Working Paper #8599). Cambridge, MA: National Bureau of Economic Research.

Harris, P. (2002). *Survey of California teachers.* Peter Harris Research Group.

Hawk, P., Coble, C. R. & Swanson, M. (1985). Certification: Does it matter? *Journal of Teacher Education, 36*(3), 13-15.

Hayasaki, E., (2003, February 10). Teacher Shortage Abates. *Los Angeles Times.* http://www.latimes.com.

Hayes, E. & Gendar, A. (2002, July 24). Pay hike lures better-qualified teachers. *New York Daily News.* http://www.nydailynews.com.

Henke, R.R., Chen, X., Geis, S., & Knepper, P. (2000). *Progress through the teacher pipeline: 1992-93 college graduates and elementary/secondary school teaching as of 1997.* NCES 2000-152. Washington, DC: National Center for Education Statistics.

HGSE News. (2003, April 22). *New research finds school hiring and support practices fall short.* Cambridge, MA: Harvard University.

Ingersoll, R. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38*(3), 499-534.

Ingersoll, R. (2002, January). *Out-of-field teaching, educational inequality, and the organization of schools: An exploratory analysis* (R-02-1). Seattle, WA: Center for the Study of Teaching and Policy.

Jelmberg, J. (1996). College-based teacher education versus state-sponsored alternative programs. *Journal of Teacher Education, 47*(1), 60-66.

Jerald, C. (2002, August). *All talk, no action. Putting an end to out-of-field teaching.* Washington, DC: The Education Trust.

Kain, J., & Singleton, K. (1996). Equality of educational opportunity revisited. *New England Economic Review (May/June)*, 87-111.

Kershaw, J. & McKean, R. (1962). *Teacher shortages and salary schedules.* New York: McGraw-Hill.

Kozol, J. (1991). *Savage inequalities.* New York: Crown Publishers.

Laczko-Kerr, I., & Berliner, D.C. (2002, September 6). The effectiveness of "Teach for America" and other under- certified teachers on student academic achievement: A case of harmful public policy." *Education Policy Analysis Archives, 10*(37). Retrieved 5/3/03 from http://epaa.asu.edu/epaa/v10n37/.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis, 24*(1), 37- 62.

Lee, V., & Burkam, D. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school.* Washington, DC: Economic Policy Institute.

Levin, H. M. (1980). Teacher certification and the economics of information. *Educational Evaluation and Policy Analysis, 2* (4), 5-18.

Linn, R.L. & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24* (1), 29-36.

Loeb, S. & Page, M. (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics, 82*(3), 393-408.

Loeb, S., Darling-Hammond, L., and Luczak, J. (forthcoming). Teacher turnover: The role of working conditions and salaries in recruiting and retaining teachers. Stanford: Stanford University School of Education.

Lutz, R. & Hutton, J. (1989). Alternative teacher certification: Its policy implications for classroom and personnel practice. *Educational Evaluation and Policy Analysis, 11*(3), 237-254.

Massachusetts Institute for Social and Economic Research (1987). *Report on the status of teacher supply and demand in Massachusetts.* Amherst, MA: Author.

McKibbin, M. (1998). Teaching internship programs: Alternative prepration and licensure in California: Purposes, procedures and performance. Sacramento, CA: California Commission on Teacher Credentialing.

Mehlman, N., *My Brief Teaching Career* (June 24, 2002, *New York Times*, need op ed page #

Miller, J., McKenna, M., & McKenna, B. (1998). A comparison of alternatively certified and traditionally prepared teachers. *Journal of Teacher Education, 49*(3), 165-176.

Mitchell, N. (1987). *Interim evaluation report of the alternative certification system (REA87-027-2).* Dallas, TX: DISD Department of Planning, Evaluation, and Testing.

Monk, D. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 12*(2), 125-142.

Mont, E., & Rees, D. (1996). The influence of classroom characteristics on high school teacher turnover. *Economic Inquiry, 34*(1), 152-167.

Murnane, R. (1983). Understanding the sources of teaching competence: Choices, skills, and the limits of training. *Teachers College Record, 84*(3), 564-589.

Murnane, R., & Olsen, R. (1990). The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina. *Journal of Human Resources, 25*(1), 106- 124.

Murnane, R., Singer, J., & Willett, J. (1989). The influences of salaries and 'opportunity costs' on teachers' career choices: Evidence from North Carolina. *Harvard Educational Review, 59*(3), 325-46.

Murnane, R., Singer, J., Willett, J., Kemple, J., & Olsen, R. (1991). *Who will teach? Policies that matter.* Cambridge, MA: Harvard University Press.

National Center for Education Statistics. (1997). *America's teachers: Profile of a profession, 1993-94.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics. (1998). *Projections of education statistics to 2007.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics. (2000). *Digest of Education Statistics, 1999.* Washington, DC: U.S. Department of

National Center for Education Statistics. (2002). *Qualifications of the public school teacher workforce: Prevalence of out-of-field teaching*

1987-88 to 1999-2000. Washington, DC: Author.

National Commission on Teaching & America's Future. (1996). *What matters most: Teaching for America's future*. New York: Author.

National Commission on Teaching & America's Future. (2003). *No Dream denied: A pledge to America's children*. NY: Author.

National Education Goals Panel. (1999). *Reading achievement state by state, 1999*. Washington, DC: U.S. Government Printing Office.

Pflaum, S. W. & Abrahamson, T. (1990). Teacher assignment, hiring, and preparation. *The Urban Review, 22*, (1), 17-31.

Pogodzinski, J. M. (2000). *The teacher shortage: Causes and recommendations for change*. San Jose: Department of Economics, San Jose State University.

Raymond, M., Fletcher, S., & Luque, J. (2001). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Stanford, CA: The Hoover Institution, Center for Research on Education Outcomes.

Recruiting New Teachers (1998). *The essential profession: A survey of public attitudes toward teaching, educational opportunity, and school reform*. Belmont, MA: Author.

Rivkin, S., Hanushek, E., and Kain, J. (2000). *Teachers, schools, and academic achievement*. (Working paper #6691, revised). Cambridge, MA: National Bureau of Economic Research.

Rothstein, R. (2002, July). Raising school standards and cutting budget: Huh? *New York Times*, July 7, 2002.

Scafidi, B., Sjoquist, D., & Stinebrickner, T. (2002, September). *The impact of wages and school characteristics on teacher mobility and retention*. Unpublished manuscript.

Schlechty, P. (1990). *Schools for the 21st century: Leadership imperatives for educational reform*. San Francisco: Jossey Bass.

Schlechty, P. & Vance, V. (1983). Recruitment, selection, and retention: The shape of the teaching force. *Elementary School Journal, 83*, 469-487.

Shields, P. M., Esch, C., Humphrey, D. C., Young, V. M., Gaston, M., & Hunt, H. (1999). *The status of the teaching profession: Research findings and policy recommendations. A report to the Teaching and California's Future Task Force*. Santa Cruz, CA: The Center for the Future of Teaching and Learning.

Shields, P.M., Humphrey, D.C., Wechsler, M.E., Riel, L.M., Tiffany-Morales, J., Woodworth, K., Young, V.M. & Price, T. (2001). *The status of the teaching profession 2001*. Santa Cruz, CA: The Center for the Future of Teaching and Learning.

Snyder, J. (2002). New Haven Unified School District: A teaching quality system for excellence and equity. In A. Hightower, M. Knapp, J. Marsh, and M. McLaughlin (Eds.). *School districts and instructional renewal* (pp. 94-110). New York: Teachers College Press.

SRI International. (2002, October). *Alternative certification: Design for a national study*. Menlo Park, CA: Author.

Strauss, R. P. & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review*, 5 (1), 41-48.

Strayhorn, C. K. (2003, January). ED 4: Alleviate the Texas teacher shortage. E-Texas: Limited Government, Unlimited Opportunity. Downloaded on 5/3/03 from http://www.window.state.tx.us/etexas2003/ed04.html.

Stoddart, Trish (1992). An alternate route to teacher certification: Preliminary findings from the Los Angeles Unified School District Intern Program. Peabody Journal of Education, 67(3).

Theobold, N. (1990). An examination of the influence of personal, professional, and school district characteristics on public school teacher retention. *Economics of Education Review, 9*(3), 241-250.

Theobold, N., & Gritz, R. (1990). The effect of school district spending priorities on the exit paths of beginning teachers leaving the district. *Economics of Education Review, 15*(1), 11-22.

Tye, B. B. & O'Brien, L. (2002). Why are experienced teachers leaving the profession? *Phi Delta Kappan, 84*(1), 24-32.

Tyson, S., Hawley, H., & McKibbin, M. (2000). *Pre-internship teaching program: A progress report to the legislature*. Sacramento, CA: California Commission on Teacher Credentialing.

U.S. Department of Education. (2002, June). *Meeting the highly qualified teachers challenge: The secretary's annual report on teacher quality*. Washington, DC: U.S. Department of Education, Office of Postsecondary Education.

Wayne, A., & Youngs, P. (in press). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*.

Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Princeton, NJ: Educational Testing Service.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10*(12); retrieved from http://epaa.asu.edu/epaa/v10n12/.

Wilson, S., Darling-Hammond, L., & Berry B. (2000). *A case of successful teaching policy: Connecticut's long-term efforts to improve teaching and learning*. Seattle, WA: Center for the Study of Teaching and Policy.

Wilson, S., Floden, R., & Ferrini-Mundy, J. (2001, February). *Teacher preparation research: Current knowledge, gaps, and recommendations* (R-01-3). Seattle, WA: Center for the Study of Teaching and Policy.

Wise, A., Darling-Hammond, L., & Berry, B. (1987). *Effective teacher selection: From recruitment to retention.* Santa Monica: Rand Corporation.

Zhou, Y. (2002, October 5). To find teachers, raise hand high and yell, 'me!' *New York Times*.

About the Authors

Linda Darling-Hammond
School of Education
Stanford University

Linda Darling-Hammond is Charles E. Ducommun Professor of Education at Stanford University School of Education. She also served as executive director of the National Commission on Teaching and America's Future which produced the 1996 widely cited blueprint for education reform: *What Matters Most: Teaching for America's Future*. Darling-Hammond's research, teaching, and policy work focus on teaching and teacher education, school restructuring, and educational equity. She has been active in the development of standards for teaching, having served as a two- term member of the National Board for Professional Teaching Standards and as chair of the Interstate New Teacher Assessment and Support Consortium (INTASC) committee that drafted model standards for licensing beginning teachers. She is author of *The Right To Learn, A License to Teach*, and *Professional Development Schools: Schools for Developing a Profession*, along with six other books and more than 200 book chapters, journal articles, and monographs on education. Dr. Darling-Hammond works on issues of education policy and practice, including school reform, authentic assessment, professional development schools and educational research. She serves as the faculty sponsor for Stanford's Teacher Education Program (STEP). As a leader in the charge for better teacher education and teacher preparedness, Dr. Darling- Hammond is instrumental in redesigning STEP to better prepare teachers to teach diverse learners in the context of challenging new subject matter standards. She also is helping to create a network of Bay Area schools of education and professional development schools (PDS) interested in working together on school reform, and learning communities for Bay Area practitioners through an ongoing series of workshops, institutes, peer coaching networks and study groups.

Gary Sykes
College of Education
Michigan State University
East Lansing, MI

Email: garys@pilot.msu.edu

Gary Sikes is a professor in the Departments of Educational Administration and Teacher Education, College of Education, Michigan State University. Recent work includes two strands. One is on the effects of school choice policies on educational systems; his most recent book, co-edited with David Plank (Choosing Choice: School Choice in International Perspective. Teachers College Press, 2003) explores this theme. The other strand of work examines the effects of policies directed at teaching and teachers. That theme is represented in a book co-edited with Linda Darling Hammond, Teaching as the Learning Profession (Jossey Bass, 1999).

Appendix A

Federal Funding for Health Professionals

Since 1944, the federal government has offered loans to students preparing for health professions careers and has supported the development of medical education programs. These programs were expanded during the 1950s by the Medical Manpower Act and in 1963 by the Health Professions Education Assistance Act, which have been amended and expanded regularly ever since. Over a half century, a strong federal role in managing the medical workforce and strengthening medical training has contributed to America's world-class system of medical training. Title 42, chapter 6A, subchapter V of the U.S. Code details the many components of this system, which includes:

1. Forgivable loans, scholarships, fellowships, and traineeships that are designed to:
    a. Increase the numbers of doctors and nurses in fields of high- demand.
    b. Improve the geographic distribution of health professionals in medically underserved and rural

   areas.
   c. Recruit as medical students individuals who are members of minority groups.
2. Investments in health professions schools, which are designed to:
   a. Underwrite the costs of planning, developing and operating training programs in specified high-need fields (currently, for example, family medicine, internal medicine, pediatrics and general dentistry), often with special consideration for projects that prepare practitioners to work with underserved populations.
   b. Create "Centers of Excellence" at specific medical schools for increasing the supply of minority medical students and faculty and improving the capacity of professionals to address minority health issues, including development of community-based health facilities.
   c. Establish area health education centers that assess regional health personnel needs and assist in the development of training programs to meet such needs, especially in underserved areas. (Some costs are funded by state and local partners).
   d. Expand training programs for public health workers, especially in "severe shortage disciplines" (e.g. epidemiology, biostatistics, environmental health, maternal and child health, public health nursing and behavioral and mental health).
3. Support for analysis concerning the health professions workforce, which aims to:
   a. Operate a uniform health professions data reporting system to collect, compile and analyze data on health professions personnel and students-in-training.
   b. Develop a non-federal analytic infrastructure (via grants to states and other institutions) to conduct research on high priority workforce questions, including projections of supply and need by specialty and location.
   c. Conduct program evaluations and assessments.

Over the years, as needs have been identified, the Congress has continued to develop innovative strategies to address emerging personnel and service needs. For example, recent amendments to the Public Health Service Act (PL 107-251) added, to existing support for health centers and the National Health Service Corps, the creation of integrated health care networks in rural areas, grants to expand telehealth resources, and expansion of training grants to include mental health professionals and increase participation of individuals in other training fields experiencing shortages, such as dentists. Using partnership strategies, some grants are directed to states to improve their capacity to recruit and distribute high-need professionals. For example, Section 340G (42 USC 256g) provides for grants to states for innovative programs "to address the dental workforce needs of designated dental health professional shortage areas in a manner that is appropriate to the states' individual needs." States may use the funds for loan forgiveness programs for dentists who agree to practice in shortage areas or who agree to provide payments on a sliding scale; for recruitment and retention efforts; for grants or no-interest loans to help dentists establish or expand practices in shortage areas; and for the establishment or expansion of dental residency programs. Through these evolving strategies and the hundreds of millions of dollars annually allocated to them, the federal government responds to local needs for health professionals and manages the labor market so that these needs can be better met.

Appendix B

Federal Funding for Education Professionals

Federal involvement in education manpower issues also emerged in the post-war era in the United States, but it has been more spotty than the steady, consistent involvement in the health professions. Rather than developing any overarching rationale or policy, federal efforts were attached to other priorities, such as national defense or civil rights, which supplied justification for a federal role. In addressing teacher recruitment needs and shortages, the national government tended to rely on incentives with limited time horizons.

The earliest legislation involved support for veterans returning from World War II. The Serviceman's Readjustment Act of 1944 contained a provision to help defray tuition and other costs for G.I.s, with teachers colleges and normal schools on the list of approved institutions. Subsequently, as the nation was drawn into the Cold War, national defense emerged as the paramount issue. Among its provisions, the National Defense Education Act of 1958 launched a loan program that became identified with its chief sponsor, Congressman Carl Perkins. Title II, the National Defense Loan Program, supplied student loans for college education, with special consideration for students with a superior academic background who expressed a desire to teach in elementary or secondary school. The program provided that any loans would be canceled, at the rate of 10% annually, for each year of service in a public school. The Higher Education Act of 1965 increased the rate of cancellation on Perkins loans from 10% to 15% for teachers who served in schools with high concentrations of students from low- income families. Such teachers also would be eligible for 100% of loan cancellation, based on extended years of service, rather than 50% available to other teachers. The 1998 reauthorization provided Perkins loan cancellations at the rate of 15% for years one and two of service, 20% for years three and four, and 30% for year five of service. The amendments added teachers of learning disabled students to those who teach in high poverty (Title I) schools or in subject-matter shortage areas, including mathematics, science, foreign languages and bilingual education, among others. These provisions remain in effect. While they are modestly helpful, these loans are a retroactive support for individuals who find their way into teaching, not a proactive recruitment device to attract college students into training programs that ensure they will be induced into shortage fields and well-prepared to teach these disciplines.

A new theme—civil rights—entered the federal mix beginning in the 1960s. In addition to the Perkins loans, the Higher Education Act of 1965 contained provisions aimed at staffing inner-city and rural schools. This act established the National Teacher Corps, which operated for the next 15 years. That program worked through grants to institutions of higher education, which were authorized to train recruits, who would serve in schools attended by poor children. Following a few months of initial training, recruits entered schools as interns on teams made up of an experienced teacher plus other recruits. Continuing their training while working, the interns received starting salaries from the districts where they worked, while experienced teachers received added compensation for team leadership. Over the years, the program was evaluated regularly and improved upon. For example, the model evolved from isolated placements in individual schools to clusters that included feeder schools to middle and high schools, and the training/program evaluation cycle was lengthened from two to five years. The act also funded fellowships that universities could allocate to support full-time graduate study at the master's level in education. A number of Master of Arts in Teaching (MAT) programs evolved out of these fellowships. These programs became, in essence, the first alternatives to traditional undergraduate teacher education. The early MAT efforts, one-year master's degree programs at places like Harvard, Stanford, Columbia's Teachers College, and Duke, later became models for many university-based alternative programs in the 1990s.

The combination of these investments in recruitment and a reduction in teacher demand led to the virtual elimination of emergency hiring of teachers by 1979. Although there were serious questions about the quality of teacher supply at that time (see e.g., Schlecty and Vance, 1983; Carnegie Forum on Education & the Economy, 1986), most federal teacher recruitment programs of the 1960s and 1970s were eliminated in 1981. By the late 1980s, however, concern about the quality and supply of teachers began to emerge again. In 1986, the Paul Douglas Teacher Scholarship Program (formerly the Congressional Teacher Scholarship Program) was authorized. Over a 10-year period until its demise in 1996, this program provided scholarships to outstanding high school graduates who planned to pursue careers in preK-12 teaching. Applicants had to be ranked in the top 10% of their high school graduating class or have GED scores in the top 10% of the state or nation. The program also operated through the states, which could add their own selection criteria in response to particular targets and needs. State criteria often included such factors as recruitment from historically under-represented groups, from low-SES backgrounds, from candidates who wanted to teach in poor schools, and for teaching mathematics and science. The program was modest in size, allocating only $15 million from 1987 through 1994. Loans under the program were forgiven at the rate of two years of teaching service for each year of scholarship award; this provision was modified to one year of teaching for one year of scholarship support for teaching in subject shortage fields. Evaluations indicated that nearly two-thirds of recipients completed teacher certification, and two-thirds of these taught.

Another program begun in 1986 sought to tap retiring military personnel for teaching. The Army began a pilot program for servicemen to enroll in teacher certification program prior to discharge. The Navy followed several years later with a program of its own. These programs worked through cooperating colleges and universities to ease the transition from the services into teaching. Some years later, the Army also established several alternative teacher certification programs for armed forces personnel, with pilots in Texas and Georgia.

The Troops to Teachers Program (TTT) began as a joint venture between the departments of Defense and Education. The 1993 National Defense Authorization Act (PL 102-484, Section 4441) formally established this program, which offered stipends of up to $5,000 to allow former members of the armed services to obtain teacher certification. In addition, school districts could receive up to $50,000 over five years for every TTT teacher they hired. Both the stipends and the grants were discontinued after 1995, but in 1999 the TTT program was reauthorized and transferred from the Defense Department to the Education Department. TTT, too, has been a very modest effort, with the 2001 appropriation reaching only $3 million, when it was placed within the Eisenhower Professional Development Program. The program operates through grants to states that submit proposals outlining the services and activities they will undertake. As of 2000, 22 states had joined the program, and 13 more were considering it. Studies that have tracked the program report high rates of participation in math, science, special education, and vocational education; more teaching at the high school level; and more teaching in the inner city. Teachers are more likely to be male (86%) and minority (33%) than the overall teacher workforce.

Under the No Child Left Behind Act, the TTT Program is a subpart of the Transitions to Teaching initiative but is still a distinct program. Participants can still receive either stipends of $5,000 a year (up to 5,000 may be awarded annually in return for a commitment to teach for three years) or bonuses of $10,000 (up to 3,000 annually in return for an agreement to teach for three years in a high-need school). The Transitions to Teaching Program authorizes 5-year grants to partnerships and eligible entities to establish programs to recruit and retain highly qualified mid-career professionals and recent college graduates as teachers for high-need schools, including recruitment via alternate-route programs that condense the period of preparation. This is a new authority in the No Child Left Behind law, but Congress provided $31 million for similar activities in 2001. These institutional funds may be used for scholarships, pre- or post- induction activities, placement initiatives, payments to schools to supply incentives for teachers, collaboration with institutions of higher education to develop recruitment programs and other strategies. Program participants must teach in a high-need school for at least three years following receipt of support.

In addition to continued funding for the Perkins loans, another part of the reauthorized Higher Education Act established the Federal Family Education Loan Program, together with a Direct Lending provision. Together, these supplied loan and principal forgiveness of up to $5,000 for Stafford loans for borrowers who agree to teach for five consecutive years in low-income elementary or secondary schools (i.e., schools where more than 30% of students are eligible for Title I aid). Loan repayment is deferred during the 5-year teaching commitment. These provisions were further amended in 2001-2002 to include three years of Stafford and Federal Supplemental loans for those who teach in a federally designated teacher-shortage area, including subject matter, grade level or geographic shortages.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State Univeristy

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil
(AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

PDF file of this article available:

This article has been retrieved 1035 times since September 24, 2003

Volume 11 Number 34 **September 24, 2003** **ISSN 1068-2341**

Constraining Elementary Teachers' Work: Dilemmas and Paradoxes Created by State Mandated Testing

Sandra Mathison
University of Louisville

Melissa Freeman
University at Albany, SUNY

Citation: Mathison, S. and Freeman, M. (2003, September 24). Constraining Elementary Teachers' Work: Dilemmas and Paradoxes Created by State Mandated Testing. *Education Policy Analysis Archives, 11*(34). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n34/.

Abstract

There are frequent reports of the challenges to teacher professionalism associated with high stakes and mandated testing (McNeil, 2000). So, we were not surprised in this year-long study of two elementary schools in upstate New York to hear teachers talk about the many ways the 4th grade tests in English Language Arts, Mathematics and Science undermine their ability to do their jobs with integrity. We came to understand in more nuanced ways the ongoing tension created by teachers' desires to be professionals, to act with integrity, and at the same time to give every child a chance to succeed. What we found in these schools is that the high stakes tests continually forced teachers to act in ways they did not think were professional and often resulted in creating instructional environments that teachers did not think were conducive to student success.

The teachers at these elementary schools are not radicals. They do not seek complete autonomy, they do not eschew the need for accountability (even bureaucratic accountability), they find some virtue in state mandated tests, they are content within centralized systems that proscribe some aspects of their work. But, they also perceive themselves as professionals with both the responsibility and capability of doing their jobs well and in the best interests of their students. New York State's outcomes based bureaucratic accountability system tests their resolve, makes them angry or frustrated, and requires unnecessary compromises in their work.

> Most of our time in fourth grade is spent test-prepping
> There is very little of the extra projects
> The extra fun kinds of activities
> That we used to be able to do
> That goes by the wayside

Because we need to test prep

Being in fourth grade is almost an advantage
If I need materials I say
Oh it's test related
Then I can get them
If I have a child that I need to have looked at
Oh it's fourth grade

There's more of an emphasis on something
Whether that's good or bad
I'm most uncomfortable at the mid-year
When it's time for us to decide
Is this child going to meet the criteria
to move on to the next grade?

You take a child to a retention committee
This child might not necessarily be ready for the next grade
But professionally I know retention is not the answer
That is no longer weighted very heavily
When you as a professional say
I know the solution for this child is not retention

What this test is testing is good
Kids should be able to read a passage
And respond to it in writing
There's nothing wrong with that
What's wrong is the way the adults in the world
Take the scores and report them

It's a benchmark
If a child can't do it in fourth grade
And they can get it in fifth grade
Why should we penalize them?
And if it takes them an extra year to master something
That's okay

We are not financial planners, where we are judged
In how many millions of dollars we brought in
We are not Wal-Mart
In how many sales we made
We are a service industry
So stop comparing success
With scores, growth, end products

What if you have a kid who got a two on the ELA
But was a knucklehead
An emotional disaster
Disruptive
But during the course of the year
In behavior
In courtesy and respect
Improved tremendously
Are you not a success then?
Did that kid not improve?

Are they measuring that?

[poetic transcription of a group interview of Willow Valley teachers]

Introduction

The current accountability strategies of school reform rely heavily on measuring outcomes, especially student achievement, and attaching consequences, either positive or negative, to various levels of performance. These accountability strategies effect everyone and every aspect of schools and schooling at local, regional, national and international levels. This article, examines the ways state mandated testing, the primary vehicle of accountability, effect teachers' work and, in particular, how their professionalism is seriously challenged by this testing.

There are frequent reports of the challenges to teacher professionalism (Note 1) associated with high stakes and mandated testing (McNeil, 2000). So, we were not surprised in this year long study of two elementary schools in upstate New York to hear teachers talk about the many ways the 4th grade tests in English Language Arts, Mathematics and Science undermine their ability to do their jobs with integrity. What we came to understand in more nuanced way is the ongoing tension created by teachers' desire to be professionals, to act with integrity, and at the same time to give every child a chance to succeed. What we found in these schools is that state mandated tests continually forced teachers to act in ways they did not think were professional, and that, in fact, this was often necessary in order to give every child an opportunity to succeed.

Context and Methodology

This year long ethnographic field study of two schools was conducted during the 2001-02 school year in two school districts in upstate New York, and is part of a larger study of the relationships among teaching, learning and state mandated testing in four upstate New York school districts. These school districts are different from many, at least at the moment, since each is participating in a National Science Foundation funded teacher enhancement project. This project is aimed specifically at providing professional development in science to elementary and middle school teachers with a pointed emphasis on helping teachers better prepare their students for the New York State 4th and 8th grade science tests. 2001-02 was the third year of this professional development project.

Our research postulates that teachers in these districts might be better able to cope with the demands of state mandated testing, certainly in science but perhaps in other subjects as well, as a result of teachers' potentially greater access to professional development. This paper does not address this issue directly, but at this stage of our research project we are doubtful that this relationship holds. This is true in part because the science tests are significantly less important to teachers, school administrators, and the New York State Education Department than are the English Language Arts and Mathematics tests at the elementary and middle school levels. Having considered this possibility, our research focuses holistically on the interactions among teaching and learning across all subject matter. Indeed, as we will discuss here the relative importance of the tests and when they are administered are key factors in decisions about curricular emphasis across the school year. Our long-term goal is to understand the complex interactions at the classroom, building and system levels among the many demands the state accountability system places on the educational enterprise.

In New York State, "outcome-based bureaucratic accountability" prevails (O'Day, 2002). This is a form of accountability that holds teachers and schools accountable to state education authorities for producing "specific levels or improvements in student learning outcomes." (p.8) These student learning outcomes are manifest in performance on state mandated tests beginning in 4th grade on through Regents Examinations required now of all students in New York's high schools. Such an outcome based bureaucratic accountability strategy focuses teachers (and students) on specific forms of limited knowledge and skills and in so doing focuses pedagogical and curricular decision-making.

The fieldwork for this study involved at least one day per week in each school--observing classrooms, talking with teachers and administrators, and attending school meetings and events. (Note 2) A great deal of our field work focused on 4th grade classrooms (since this is where the testing burden primarily lies) but we observed classrooms and talked with teachers at every grade level. Additionally, a focus group interview with teachers and a focus group interview with parents were conducted, as were individual interviews with building and district administrators. Throughout the data analysis, we engaged a number of teachers and the principal at each school as peer debriefers, continually checking our understandings and reading our case studies.

Table 1 summarizes descriptive information about the schools and districts and Table 2 indicates the schools' pass rates on the ELA, mathematics, and science tests for the past three years. Table 3 illustrates the range of state mandated tests given in New York state elementary schools. Included in this table are the dates the tests are administered and the format. Both are critical elements in teachers' decisions about what to teach, how, and when. Additionally, but not part of this study, New York has adopted, under the leadership of Commissioner Richard Mills, the "Regents for All" Plan which will require *all* students pass a minimum number of courses and Regent's Examinations in five subjects to receive a State recognized high school diploma.

Table 1 Description of Schools and Districts

School	District	# of Students	# of teachers	Free/reduced lunch	Race/ethnicity	English Lang Learners	Grade levels
Hemlock Elementary*	17 buildings; urban; overall 69% of students are on free/reduced lunch; drop out rate 7%; 9000 students	395	30	90%	52% white 35% Black 12% Hispanic 1% other	0%	PreK - 5
Willow Valley Elementary*	2 buildings (elementary & ms/hs); working class, predominately white; 1500 students	818	52	46%	93% white 5% Black 1% Hispanic 1% other	2%	K - 6

Source: 2002 New York State School Report Cards

* pseudonyms are used for schools

Table 2 Test Scores (% of students "passing" 4th Grade State Tests)

School	Year	ELA	Math	Science
Hemlock*	1998-99	15%	48%	
	1999-00	40%	53%	62%
	2000-01	50%	63%	63%
Willow Valley**	1998-99	44%	71%	
	1999-00	48%	72%	56%
	2000-01	55%	77%	77%

*This school did not meet the state standard in ELA, but made adequate yearly progress (AYP) in 2000-01.

**This school met the state standard and made adequate yearly progress (AYP) in 2000-01.

Table 3 New York State Mandated Elementary Tests (2001-02)

Grade Level	Fall	Spring	Test Format
4th		English Language Arts (early Feb)	Reading & 28 mc questions Listening & written responses Reading & written responses Independent writing prompt
		Mathematics (early May)	30 mc questions Short and extended responses
		Science (May)	45 mc questions Performance--5 stations; 4 questions/station
5th	Social Studies (Nov)		45 mc questions 3-4 constructed responses 1 document based question

Contexts for Teachers' Dilemmas

Teachers may never have had much autonomy and the professional status of teaching cannot be taken for granted. Teachers' work has historically received low pay, been perceived as relatively low status, and often operates within authoritarian and often petty school cultures (Katz, 1971). "Education has not suffered from any freedom granted teachers to run schools as they see fit; it has suffered from the suffocating atmosphere in which teachers have had to work" (p.131). Still, much educational research demonstrates the centrality of teachers in educational reform (Elmore, 1996), they are "curricular-instructional gatekeepers" (Thornton, 1991). Schools have also been the locus of almost every social change effort placing ever more demands on teachers (e.g., drug education, sex education, values education, environmentalism, bus duty, data management) with no reprieve from prior demands. The current standards based reform movement with its clear specification of content, pedagogy, and assessments adds to these demands, increases authoritarianism, and further erodes teachers' sense of professionalism (Madaus, 1998; Mathison, 1991; Noble & Smith, 1994; Ross, 2000; Vinson, Gibson & Ross, 2001). In a study of Kentucky teachers after the implementation of the Kentucky Education Reform Act, Kannapel, Coe, Aagaard, Moore & Reeves (2000) conclude that, "the educators we spoke with resented the accountability measures as an insult to their professionalism."

There is ample research describing how state mandated tests, particularly high stakes tests, challenge and compromise the professionalism of teachers. McNeil's (2000) research in Texas illustrates a range of constraints on teachers' work, constraints that lead them to "exclude their richest knowledge from their lessons" (p.192). These constraints spring from the increased standardization and specification of important knowledge as that which is on the test. As a result, teachers adopt generic forms of content and presentation; develop a "test based curriculum"; separate content "for the test" and "real content"; further fragment knowledge; and even retire. Testing leaves little time for "real instruction" (Hoffman, Assaf & Paris, 2001). In some cases, when a mandated test demands something that has not previously been a routine part of the curriculum, such as writing or problem solving, there is refocusing although in ways driven pointedly by the test (Hillocks, 2002; Kannapel, et.al, 2000).

Teachers do not feel good about the constraints that testing places on their work. McNeil (2000) describes teachers moving away from particularized child centered teaching to teacher centered generic teaching, because the latter reflected state mandated curriculum and assessments. Dramatically, she concludes: "The reforms required that they choose between their personal survival in the system or their students' education" (p.192).

The schools in this study reflect findings of other researchers. Teachers at Hemlock and Willow Valley Elementary Schools perceived their professionalism to be diminished. Through outcomes-based bureaucratic accountability teachers' work has come to be defined by the state-mandated tests, especially in English Language Arts, as well as district directives geared to improve state test scores. But for these teachers it is not an either or choice between personal survival and the students' education. These teachers confront the dilemma of being a good teacher, a professional, and helping kids to succeed, which is marked by performance on state tests. What we saw repeatedly was that this dilemma is almost always solved in favor of the students, that teachers sacrifice their professional integrity in order to help every child be as successful as s/he can be on the tests, even when they lack faith in the indicator. This resolution plays itself out in the classroom as well as around the administration and scoring of the state tests. The following sections elaborate how teachers experience and come to uneasy resolutions of the dilemmas they face.

Faith in Children

The popular media and politicians often portray teachers as contributing to the low achievement of children, especially children of color, by having low expectations and lacking the faith that all children can learn. The political slogan, "No Child Left Behind," which titles the current Elementary and Secondary Education Act is a manifestation of this belief. However, the teachers in these schools, both in word and deed, challenge this representation although like teachers everywhere they talked of the overwhelming social forces on children's lives outside the school building. And, they did not always feel they were able to compensate for a lack of experiences (such as rich early literacy experiences) or life circumstances (poverty, violence, homelessness).

This was especially true at Hemlock Elementary, a school where most children are on free or reduced lunch and many are African American. "These are not children that don't learn. These are children that do learn--slowly." "We are being judged on something that is largely out of our control," Hemlock teachers explain as they relate stories of student absenteeism, high mobility, and academic need. "And what does it do to the individual kid? If we have a child who's a slow learner, that is a huge concern that is being left out of this testing thing by the media and politicians and the Regents. They don't want to know that there is such a thing as a slow learner. And to tell a child, who gets to this higher level in a school year that they are a failure because they didn't reach this goal is horribly wrong, horribly wrong for that child."

There is less confidence in the children and teachers have a more limited sense of efficacy at Willow Valley. Willow Valley Elementary is a huge school, a consolidation of three elementary buildings into one, occupying an office building complex the school district acquired from a downsized business. Students here are white, working class and poor, and living in a neighborhood enclave cordoned off by industry and freeways. Teachers here frequently characterize the school's students as a high needs population: "It's hard and with the special ed kids… we need consistency and structure. As soon as the tiniest, tiniest thing changes, they're very needy in that sense. That so terrifies me about them going into fourth grade because their independence to be able to, even on a very simple task, read the directions and complete it… As long as everything is being modeled step-by-step or very guided or very structured, they're fine. But as soon as you look for that independence, they struggle." Parents are aware of the characterizations of their children: "Labels [that they give our kids]--you go to the school board meeting and you hear this, go to a PTA meeting, go to a committee meeting, and it's the socio-economic background, it's the transient populations. So, because of this we can't expect a good education for our children?" The principal is aware of the strong tendency to view high needs students as somehow less able than others and feels it is his role to continuously stress that teachers need to learn to work with what the students bring with them, not what they aren't bringing.

Willow Valley teachers don't give up on children, but they often express reaching the limits of their capabilities. "We are doing what good 4th grade teaches are supposed to do, we're teaching the students the curriculum. You can't ask us to make up for the fact that this child is deficient in this skill and has been since kindergarten. There are just, I don't know how to describe it, there are just certain things that are beyond the 4th grade classroom teacher's control and yet we are being asked what are we going to do about this child? I can't do anything more. I've done everything I can do. You have to pass it off to somebody else now." But still, teachers worry about what will happen to children, "it still eats us," and repeatedly we saw teachers making school instructive and enjoyable for their students.

In the classroom

Teaching to the test

The many meanings of 'teaching to the test' and the validity of the test itself conspire to create anxiety about the right thing to do. The basic tenet seems to be: if a test measures what is important then teaching to the test is okay, but if the test is misdirected or poorly constructed or only a partial picture of what is important then teaching to the test is not okay (Heubert & Hauser, 1999; Smith, 1991). The difficulty for teachers is that they often hold both views simultaneously. The 4th grade ELA encourages them to teach more writing than they have before and the 4th grade Math Test encourages them to teach more problem solving--so teaching to the test (in the sense of taking curricular cues from the test content) is good. But, the reading and writing on the 4th grade ELA is formulaic and focuses on syntax, and discourages creativity, exploration of language, and discussion--so teaching to the test is bad. Coupled with a context that defines these tests as high stakes tests, especially the ELA, with serious consequences for schools (threats of state intervention), for teachers (shame and rewards), and students (possibilities of retention in grade, labeling), and teachers are left with little choice. They teach to the test. At Hemlock this is a highly structured, orchestrated effort while at Willow Valley this is a more haphazard, individual response.

Content

> "I'm finding that I used to read stories for enjoyment. And now when I'm reading a story I'm trying to think, 'Alright, now how am I going to use this?' And I'm trying to get the contrast and compare. And trying to do author studies. And I almost find that I'm not enjoying it. I'm enjoying it, but it's not like it used to be when we could read a story put it aside and maybe do a tracing and cutting activity to go with that story. I'm not doing so much cutting anymore. I'm doing a lot more, I'm trying to do critical thinking and we're writing in journals. It's not a fun thing anymore. I'm trying to always get two jobs done as one. How can I use this twice? How can I really push this? [Willow Valley teacher]

Teachers value what they perceive to be positive changes the tests have instigated in their teaching, resent what they have had to give up to make these changes, and sometimes defiantly teach what they think is important even though it may not help the students do well on the test.

At both Hemlock and Willow Valley schools teachers believe the state mandated tests have changed what they teach, and often for the better. Teachers believe the ELA "is a good test. It tests listening, reading and writing at an appropriately high level." The Hemlock reading teacher describes the test as focusing on "higher order thinking skills, therefore in our program the emphasis has been changed from the lower level thinking skills such as recall and detail to the higher level skills. That's a benefit. Another benefit is that we focus on writing much earlier… due to the nature of the test we've gone from filling in the missing word, which is a former emphasis, to understanding main idea, inference, conclusions, predicting and those are all higher level skills. So the result for the students is that they are getting really a much higher level instruction now than they used to." And a 3rd grade Willow Valley teacher now does, "a lot more note taking, lots of graphic organizers, and I don't think if it wasn't for the test that I would use them in such detail." In math she teaches the concepts and skills the 4th grade teachers say the kids need, but she goes on, "I feel like I'm very much rushing to say 'we've covered it and they've at least seen it' but not giving them the practice they need." Teachers identify positive changes in their curriculum because of the ELA and math tests, but seldom mention the social studies or science tests.

Recognizing the ELA test required more of their students than they had expected in the past, the Hemlock teachers used Title I money to develop a curricular strategy to prepare their students to do as well as possible on the ELA test. "We spent a lot of time analyzing tests. At the same time we were making a huge effort to integrate. We were choosing materials and making selections do double duty with science or social studies, working around the themes so there was a whole integrated package." Teachers used trade children's literature magazines (*Ladybug* in 3rd grade and *Spider* in 4th grade) as the texts and developed multiple choice and short answer questions (like those on the ELA) for each story. As the teachers were developing this trade magazine based curriculum, the district curriculum committee adopted a basal reading series (Scott-Foresman's Reading) that Hemlock teachers are required to use. Language arts instruction now consists of the regular classroom teachers teaching the basal reading series while the reading teachers travel from classroom to classroom armed with magazines and packets of ELA test-like questions providing fast paced, no nonsense instruction of material that resembles that on the test.

Teachers at Hemlock think adoption of the basal readers is an insult and a distraction. In conjunction with the text book adoption (in both language arts and math) are messages from the district office that all teachers should be on the same page at the same time. "A lot of teacher hours went into the curriculum that they produced and then when we got our new reading series it was imposed on us… it is a mandate that you're on a certain page in a certain week across the district [and this] is unrealistic depending on the kids' abilities. So the teachers here just feel like all we're doing is frustrating our children. We are not teaching them the way that we as professionals should be allowed to help all of our children learn." The teachers have more confidence in their own trade magazine based curriculum to prepare the students.

Even though teachers feel the tests, especially the ELA, has challenged them to teach more and better they stick very closely to the forms of knowledge on the test. And so there is a question about whether students are engaged in higher order thinking or merely the appearance of such. The ELA and math tests are scored as a 1 (serious academic deficiencies), 2 (needs extra help), 3 (meets the standards) and 4 (exceeds the standards) and these levels have become an organizing structure for teaching. In fact, some form of this scoring rubric is posted in every classroom in both of these schools. This excerpt from a 4th grade classroom observation illustrates how being pushed by the test to have higher expectations is simultaneously dulled by the test.

> This class is reading *Velveteen Rabbit*. The teacher passes out a worksheet and tells the students she is going to give a response that is a 4, or a 3, or a 2. She directs them to put a 4 on the back of their worksheet and an arrow next to it.
>
> T: If I'm going to write an answer that is going to score a 4, what does it need?
> S: Answer complete.
> S: Neat.
> T: I agree, but I wouldn't worry about neatness first.
> S: Topic sentence.
> T: Yes, you need to have some sort of topic sentence. You need to remember to restate the question. What else?
> S: Details.
> T: YES, details, details, details. Where do you get the details?
> S: In the book.
> T: Ok, it's complete and it has a topic sentence. What else will people scoring be looking for?
> She reminds them about the 'Daily Language Activity' hints she gave them in the morning--punctuation, spelling, capital letters, and correct grammar.
> T: Leave a space and put a 3. What is going to be the difference between a 4 and a 3?
> S: One of those things is not included.
> T: Everything needs to be there. It will be mostly complete. Will it be perfect?
> S: No.
> This lesson continues until they have gone through 4, 3, 2, 1 and then the teacher shares some examples of responses to the question, "Why does the Velveteen Rabbit feel plain and ordinary?"
> T writes: "He feels plain." The students give it a 1 because it is too short. The teacher comments that we don't know who 'he' is and comments on the need for more details.
> T writes: "The Velveteen Rabbit feels plain and ordinary." The students give it a 3. The teacher disagrees and gives it a 2. She says it is missing details from the story--have you proven it from the story?
> T writes: "The Velveteen Rabbit feels plain and ordinary because all of the toys make fun of him. For example, the expensive toys snub him and make him feel commonplace." The teacher tells them this response is a 4. One girl copies the answer but pauses to say she disagrees, that not all the toys make fun of him because one doesn't. The teacher agrees and changes the word all to most.

In spite of the pressures of the tests, teachers do exercise their professional judgment, almost with an air of defiance, and do what they think is right by the children even though it isn't consistent with the district's curricular mandates and may not be directly tied to the test. These acts of defiance are frequently tied to helping children feel successful, encouraging them, giving them an opportunity to have fun. One district uses *Everyday Math* and this teacher describes "absolutely breaking the rules of *Everyday Math*." "All of the [students] failed the multiplication test. They didn't know how to do the partial products algorithm. They felt stupid, they felt incompetent, and they failed it miserably because their brains couldn't process all those steps at one time. So I've gone back now, I've spent two class days teaching them, doing a task analysis first, which comes pretty naturally after you've taught the multiplication algorithms. I carefully added each step, if you skip one of those steps kids like this will not be able to make that mental jump, they can't do it, you have to go in a methodical way, they have to master each step, and then they feel good about themselves. They were begging me for harder problems. They get turned on by that. They love it. Now they're going to go home, they 're going to do this homework they made up and they are all going to know how to do partial products algorithms, which I guarantee will be on the test." But she adds weakly, "Not partial products, but multiplication problems."

These teachers struggle with the fear of falling behind in a system that frowns on those who do. Instead of comfortably working on what they perceive their students need to better understand the material, they push ahead until it is obvious that pushing ahead is causing their students to fall further behind. The curricular calendar and the testing schedule do not stop for make-up time and so the pressure is to catch up by covering material superficially.

Textbook Adoption

District textbook adoption occurred in both of these districts as a result of the state standards and tests. And textbooks are chosen to match the tests, not a difficult thing to do given that the textbook and test publishers are often one and the same. While these new textbook adoptions filled a void where there previously had been few resources, they also create chaos and conflict. In the case of Hemlock, the adoption of a basal reader diverted teachers from a curriculum they had created. At Willow Valley, some teachers did find the time to do "double entry teaching." "What I end up doing is double teaching because I'm teaching the series and I'm also teaching using the strategies and the plans that I had when I taught novels. I'm basically double dipping for them, but you have to in order for them to get all of the skills. And I can't teach skills in isolation. What good is teaching them the "short a" sound in ten words if they are not going to use it within a story and be able to read it. You look for stories like *Little Bear* that would have that "short a" sound within it, so now they can apply the skill they learned."

The other consequence of district wide textbook adoptions is a perceived added difficulty in integrating the curriculum. Because time is a scarce commodity, and teachers understand the priority of language arts, they would like a curriculum that provides language arts skills through math, science and social studies content. The Hemlock teachers had selected trade magazine stories with science and social studies content for precisely this reason. They now have textbooks that are a giant step backward in terms of integration. "As happy as I am to have a standardized curriculum across the district, this new reading program has no fourth grade social studies content and no fourth grade science content. None."

In many ways, these teachers are faced with a richness of resources but lack the time, guidance, and support for creating an integrated curricular whole out of the textbooks, trade materials, math series, science kits, newspapers, test preparation materials. One teacher summed up this frustration, "You have to wonder, do you do the math in the reading series or the reading in the math series?"

Pedagogy

The Hemlock Elementary plan to better prepare their students also dealt with how language arts would be taught and incorporated more ELA focused instruction by reading teachers in all 3rd and 4th grade classrooms. The ELA curriculum included blocking off specific times in each week at each grade, breaking students into four homogeneous groups and having four teachers working with each group in a different spot in the building. Groups were based on Terra Nova test scores, teacher judgments of reading ability and students' potential performance on the ELA tests--as solid 3s, 3s but potential 4s, 2s but potential 3s, and 1s and 2s. Teachers are confident that small homogeneous groups working closely with a teacher is the best way to meet the students' individual needs and capitalize on their strengths. "[Teachers] who had the higher groups could do a lot more of the advanced higher order thinking skills, whereas my kids would be doing a lot more of the decoding, word recognition and basic lower level comprehension skills."

This plan was thwarted by the superintendent who decreed that children could no longer be pulled out or grouped in preparation for the ELA test, and this decree left teachers feeling betrayed, undermined. The district is attempting to promote inclusion and to disrupt a tracking system that takes root in the early years of schooling. The district response was totally unexpected and seems illogical to the teachers--they are still permitted to group and use pullout strategies in math. The school's ELA scores had gone up dramatically with the teachers' plan and they have profound confidence in the power of grouping and pull out strategies. Expecting recognition, the blow is huge. "Now I wouldn't dare pull a student out to help them improve. We were told in no uncertain terms that we had to follow policy. The removal of the principal [because she permitted teachers to use this strategy] was a message to staff. First, we got the news of how well we had done. We were shocked and ecstatic, and then totally demoralized. We were stunned." Whether grouping and pull out programs are a good or bad idea the dynamics here suggest an undermining of teacher professionalism even though all parties are driven by an effort to help the kids do well on the indicator that matters most, the ELA test.

The Hemlock strategy of dedicating the reading teacher to do the "ELA curriculum" and the classroom teacher to teach the basal reader created additional challenges to teachers sense of being a good teacher. New teachers are especially frustrated: "We don't decide what is taught during that time. It's all reading teacher." Teachers' professionalism is compromised in two ways by this test score improvement strategy. First, classroom teachers are left standing around watching while reading teachers use direct instruction techniques (which some do not agree with) thus wasting valuable resources that could be used to help children. Second, this strategy leaves teachers in a bind if the reading teacher is absent or late. Sometimes they find themselves singing songs or having students read quietly, not wanting to start something new until they know what is going on. And if the reading teacher does not show up they do not have the ELA materials and have to substitute other content. On one such occasion the teacher remarked that he had been promising the kids they would do social studies and the absence of the reading teacher is what made that possible.

District textbook adoption, common curricula, standardization weigh heavily on teachers, challenging the fundamental notions of individualizing education, child centered teaching. Teachers acknowledge they need to measure students' reading, comprehension, and so on but feel they are caught on the horns of a dilemma of standardization and individualization. They are forced to ignore individual strengths and needs in an attempt to get all children ready to tackle the same test at the same time. "There are deep contradictions in the messages we are getting. Every kid is supposed to have and indeed we are supposed to encourage them to build on their individualized learning styles. The district actively supports individualized educational programs for children and then we are supposed to cram them through the test using the same approach for all children. Give me a break!"

Splitting the Curriculum

McNeil (2000) describes teachers' use of "double entry lessons" that split the curriculum into the real content and the official (tested) content. Such a strategy would be seen as a luxury by the teachers at Hemlock and Willow Valley where time is a scarce commodity and teaching the official (tested) content takes all the time there is, and more. The strategy that has evolved in these schools is a splitting of the curriculum according to the relative importance of the test and the time of year the test is administered. Although there are 4 tests given at the elementary level in New York, everyone implicitly understands that the ELA is what matters. Reading and language arts are seen as the basis for all other subjects (and, in fact, a common criticism of all other tests is that they test reading as much as science or math or social studies) and so take precedence. It is the ELA scores that have been used for decisions about remediation, retention in grade, teacher quality. Table 3 indicates when each test is administered in 4th grade--ELA in early February, followed by math and then science in the spring. So, in primary grades and especially 4th grade the school curriculum is language arts intensive until February, followed by a couple of months of concentration on math, and much more limited emphasis on science. And, 5th grade teachers should not expect that students will be prepared during 4th grade for the social studies test which is given in November of the following year for a 4th grade cohort of students--there simply is no time.

"We structure our whole day in 4th grade right up through January, our whole day is structured towards the ELA, and then after that, after the ELA, there will be a shift in focus and then we will be structuring our entire day to focus on math and science." About 4 hours each day from September to January, the teachers prepare students specifically for the three days of ELA testing, for the moment in time when teaching and learning stop, when Hemlock stands still for the test. And the same rhythm repeats itself at Willow Valley Elementary. "So I find that I often put social studies and science on the back burner to get through the reading and the writing. And I find that I'm spending a good 2 1/2 to 3 hours a day on language arts and I'd rather not. I'd rather be able to teach every subject every day and that doesn't often happen in my class. I wish it did, but it doesn't. Right now we are under the gun, we are under pressure. You hear it from the administration, you hear it from colleagues, "Do you think they are ready?" and they don't do it to nag you, it's a concern." Another teacher anthropomorphizes science: "Poor science--it's really been pushed aside. How am I going to get [the students] ready for the science test in two weeks?"

Two days after the ELA test at Hemlock, the teachers are smiling; the pace is more relaxed, the discipline looser. In a 4th grade class, students are tackling a deductive reasoning problem. They are given clues and use them to deduce the correct answer. The lesson is interactive. There is talking among the students, and questioning and sharing between students and teacher. The students are engaged and interested. This is a welcome respite before serious preparation for the state math test begins.

These classrooms are unlike our traditional images of elementary school classrooms that focus on language arts, especially reading, in the morning while children are fresh and attentive, and then move to mathematics and finally science and social studies in the afternoon, with special subjects interspersed throughout the week. Because of the testing, the curriculum has been split across the school year, not across the school day or week. And, although language arts has always consumed most of the time in elementary classrooms, it is even more so in these schools.

The Test, Itself

During testing

When the tests arrive at schools the tension rises. Teachers must watch their students take these tests and adhere to New York State Education Department instructions about test administration. Sorting through how to administer the test, what questions can the teacher answer, how should the accommodations for special education students be implemented is a dance the teachers do throughout the testing. And, while teachers are mindful of following the rules they interpret the directions differently. Some teachers are adamant about not answering any questions and watch in silence as some students struggle, others simply sit, and many work diligently on the test. Others encourage students to ask questions hoping they will be ones teachers can answer: "Today when you are doing your questions, get your hand up and ask. Most of the time we could answer your question."

During the days of a test, teachers do quick checks on student scores, analyze the test questions, check up on students, talk with them about their perceptions, give them moral support, reprimands, and teach cram sessions based on the teacher's preview of the test. In one 4th grade class after the first session of the math test, the teacher asks two boys, "How was it?" The students respond, "easy" "fun" "boring." And then two boys ask the teacher if 50 ´ 50 = 250. She has them figure it out and they find the answer is 2500. She shows them another way to solve the equation. The teacher laughs, grateful that the boys thought the test was easy, oblivious to the fact "they have no clue." And she goes on, "Are they trying to use something I taught, then that's important to me, not so much that they got it right." In another classroom just before the second day of the math test, the teacher is more focused. She hands pencils to students that say "4th graders are #1" and tells them, "These are special pencils that only work on this portion of the test." But before they begin the test she gives the students a quick refresher on parallel lines, perpendicular lines, trapezoid, parallelogram, hexagon. And she makes a last minute plea that they remember what they have learned about probability and fractions. As the students take a bathroom break, this teacher looks over the test and her mood sinks noticeably--too many factions, decimals, but then a sigh of relief, a graphing problem. "We've done at least 5 of these in our graphing unit."

In another 4th grade class after the first day of ELA testing, students color, play board games, and play on the computers while the teachers gather the tests and make charts and record student scores. Teachers compare notes on how hard they felt the test was and how well their students did. Question by question, teachers analyze the test. One teacher does an item difficulty analysis. With this information they hope they will be better prepared next year.

In another class, after weeks of intense preparation for the ELA, a teacher watches silently as her students finish the second day of the test. Once the test booklets are collected she tells the students to sit down and listen because she is going to yell at them. And she does. "I know that was a long test. But I cannot believe--I was ready to scream when I saw you sitting there staring into space. Don't tell me you couldn't have found one run-on sentence, a spelling mistake, or checking bullets against your answers to make sure you covered everything. Two half-hour sessions is not too much to ask of a 4th grader. We've worked all year on this. You can put 10 minutes more effort. Please tomorrow, don't just sit there. Find something to fix. I saw someone spell first, f-r-s-t. If I go to read them and I find that I was wrong, I'll take it all back. But if I find that I am right, I'll be even madder than I am now. Tomorrow you have another writing session. Only tomorrow you will use all your time."

Teachers know these testing moments cannot judge the quality of their work, but they find themselves acting as if this were so. And sometimes acting in ways that may not make them proud of themselves as teachers. "I have to come to balance in my own head, about how to keep the kids just as short of being over the line with stress themselves. They are children, they have to play and have fun. They are nine."

Scoring the tests

Schools in New York are responsible for scoring their own state tests. A number of teachers indicated that scoring the tests is a critical experience for understanding the content of the test and what constitutes a 4, 3, 2, and 1 response. (This experience has been important in the past because all elementary and intermediate tests were secure, but beginning with 2002 schools may keep the tests and use them to prepare for the upcoming year's test.) It takes a small group of teachers a full day's work to score any given test, a hidden cost of the state's accountability system. New York State Education Department provides training videos and materials to be used in every scoring situation, and the scoring session begins with a review of the rubrics then scoring a sample of responses. Once they begin scoring teachers discuss disagreements or questions.

During the math test scoring the questions and concerns teachers have stem from an interest in being fair to the student. In this session the first issue that arises is around responses that give an answer but do not show any work. The rubric clearly indicates that students should receive NO points if they answer the question correctly but do not show their work when it is required. One teacher sees that the student did the work, but erased it. If you can still see it, does it count as shown work? The teachers agreed that it does. And the discussion among the teachers and the facilitators deviates from the rubric and resolves the meaning of "shown" work. And the resolution favors students on both counts.

T: Answer correct, but no work?
F: Give partial credit.
T: But what if the work is there, only erased?
F: Full credit, as long as you can see it.

The next issue to arise is in scoring a graphing problem--students can get a 3, 2, or 1. A teacher asks about the meaning of a 3 score, which the rubric says is a complete and correct answer, and a 2 which are given if some information is missing. The answer that sparks the discussion is a student's graph that is complete but for the exception of one unlabeled axis. There is a title, one axis is labeled and numbered, the names for each bar are given (e.g. horses) but the axis label (e.g. animals) is missing. "Obviously he knows how to make a graph, why should he be penalized? Does he have a complete understanding of what goes into a bar graph? Yes." Another teacher sympathizes, "We had that problem in the past and we had to give them a 2." But the teacher is not mollified and his fellow scorer says, "If you feel so strongly about it, do what you want to and give it a 3. If you go by what [the State's rubric] says, give it a 2." The facilitator intervenes, trying to calm the outraged teacher and eventually he gives the student a 2 and turns to the next student response to find exactly the same scenario. But this time the student gets a 2 because they had all the correct labels even though the bars in the graph were incorrect. "This child obviously did not understand the concept of making a graph but because she was able to follow the directions and knew enough to label, she gets the same points as the other who obviously understands how to make a graph but forgets one label. That's not right." Much like teachers redirect students to focus on preparing for taking the test, this teacher is redirected to get on with the scoring.

F: That's why you can't compare answer to answer. You have to go by the rubric.
T: OK, then you can't compare scores. You can compare scores between schools, yet we can't compare one answer to another? You're telling me that that child has the same comprehension as the other one? Right now I could fight with the state!
F: Stay on task, we have only an hour.
Another teacher interjects with a new question,
T2: If answers are completely wrong, but the process is correct?
F: It's a partial--1.

Scoring the tests leads teachers to question their judgment, the judgment of others and especially the possibility that they may have scored too harshly. These moments of uncertainty arise especially when scoring items that require students to show their work or write an explanation. Teachers agonize over finding something salvageable even in the most incomplete answers. Again, while scoring the math test teachers have to work through what it means for a student to show 'at least the beginning of a process.' The New York State Education Department help line provides them with no guidance and they conclude:

F: If we can defend our score and our interpretations then let's do it. We can give credit for the start of a correct process if it ultimately leads to the correct answer.
T: When in doubt err on the side of the student.

With this exchange, it became clear how to resolve many uncertainties--when in doubt err on the side of the student. And this is what the teachers did and the scenario repeated itself when teachers scored the science test although always with much discussion. Interestingly, this is an issue that is specifically addressed in an informational Q and A memo from New York State Education Department that says:

> Q: On borderline calls, when deciding between adjacent score points, should the scorer always give the "benefit of the doubt" to the student and award the higher score?
>
> A: No. Such a practice can result in scoring "drift." After scoring a number of responses, a scorer may gradually, even unconsciously, begin to accept less (or demand more) than is appropriate in awarding a particular score point. Scoring "drift" can create an unfair situation where a student response could receive a different score from the same scorer depending on when the response was scored. To prevent "drift" and maintain the consistency and accuracy of all scores, it is helpful to refer occasionally to the student responses used in the training materials as examples of the various score points. These responses are often called "anchor papers" because they help to fix the acceptable range within a score point and prevent the scorer from "drifting" higher or lower in their expectations for awarding a score point. Scorers should also be encouraged to consult their Table Facilitators and Scoring Leaders with responses that seem on the line between two score points.

Even at this last moment, when teachers can help students be as successful as they possibly can be on the state tests, they do so. They follow the rubric as well as they can because they believe a great deal of effort has gone into creating them, but they are willing to "give the student the benefit of the doubt."

Conclusions

The teachers of Hemlock and Willow Valley are forced into untenable situations fraught with dilemmas that are difficult to resolve and maintain teacher professionalism and help all children to succeed to the best of their ability. Repeatedly we saw teachers put in lose-lose situations. They act in ways that are inconsistent with what they believe to be best teaching practice in order to increase the likelihood that students will succeed as measured by the state tests, which at least for many teachers is a poor indicator of the achievement and success of children. Teachers must often do the wrong thing in order to do the right thing, sort of.

It is essentially a utilitarian ethic that underlies test driven curricular reform, one based on means--ends arguments (Mathison, 1991). The New York State Education Department adopts the view that the ends justify the means, and teachers too are drawn into this logic. The means are approaches to teaching and content that teachers might not chose--that do not represent good professional practice and, the state's desired ends (high test scores) are a poor but powerful proxy for the teachers' desired ends (the contextually appropriate success of every child).

The experiences of these two schools tell us a great deal about the impact of state mandated, high-stakes testing and this paper has specifically focused on how these tests challenge teachers' professionalism, especially with regard to how they treat children. Of course, this is an interesting argument only if these things matter. These teachers wonder if policy makers and politicians have any sense of children's individual differences and the centrality of that concept to teaching *and* learning. Current state standards based reform and assessment policies and practices would suggest that policy makers and practitioners either have no sense of this, or maybe they don't care, or maybe they are trying to redefine these ideas. Through the currently proffered solutions to problems of education, policymakers/politicians/corporate CEOs eschew what teachers know about human learning and cognition, and much of what teachers know is helpful and harmful to children's achievement.

Are policy makers and politicians unaware that outcome based bureaucratic accountability driven by state mandated tests will reduce teacher professionalism and autonomy? That some research (see O'Day, 2002) suggests lower performing schools will actually lose ground? And that these accountability strategies do relatively little to alter the fundamental injustices in schools and society, such as racism and classism? We don't know for sure, but we think probably not. There is a fundamental disagreement about what kind of work teachers and students should be doing in schools--work that requires real critical thinking that may contribute to the evolution of a just and equitable society or work that has the appearance of critical thinking and will contribute to oppression. (These are not simple political disagreements; they are disagreements connected with power and money. For a more detailed discussion of this argument see, Mathison, Vinson & Ross, 2001; Vinson, 1999.)

> "By insisting that legitimate learning necessarily presents itself in and on the basis of test scores, such testing refuses to admit and accept differences (individual as well as cultural) in knowledges, values, experiences, learning styles, economic resources, and access to those dominant academic artifacts that ultimately contribute to both the appearance of achievement and the status of cultural hegemony upon which standards-based reforms depend. In effect, standardized testing encourages a singular and homogeneous public schooling—one antithetical to such contemporary ideals as diversity, multiculturalism, difference, and liberation—vis-à-vis an underlying and insidious mechanism or technology of oppression, one in which the interests of society's most powerful (the minority) are privileged at the expense of those of the less powerful (the majority)" (Vinson, Gibson & Ross, 2001).

The teachers at Hemlock and Willow Elementary Schools are not radicals. They do not seek complete autonomy, they do not challenge the need for accountability (even bureaucratic accountability), they find some virtue in state mandated tests, they are content within centralized systems that proscribe many aspects of their work. But, they also perceive themselves as professionals with both the responsibility and capability of doing their jobs well and in the best interests of their students. New York State's outcomes based bureaucratic accountability tests their resolve, makes them angry, and requires unnecessary compromises in their work. These teachers are more angry or frustrated than better, and with little indication that student achievement is advancing in genuine ways or that schools are being reformed.

References

Darling-Hammond, L. (1990). Teacher professionalism: Why and how? In A. Lieberman (Ed.), *Schools as collaborative cultures: Creating the future now.* (pp.25-50). Bristol, PA: Falmer Press.

Elmore, R. F. (1996). Getting to scale with successful educational practices. In S. Furhman & J. A. O'Day (Eds.), *Rewards and reform: Creating educational incentives that work.* (pp.294-329). San Francisco: Jossey-Bass.

Heubert, J. P. & Hauser, R. M. (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington, D.C.: National Academy Press.

Hillocks, Jr., G. (2002). *The testing trap: How state writing assessments control learning.* New York: Teachers College Press.

Hoffman, J. V., Assaf, L. C. & Paris, S. G. (2001). High stakes testing in reading: Today in Texas, tomorrow? *Reading Teacher, 54*(5), 482-92.

Kannapel, P. J., Coe, P., Aagard, L., Moore, B. D. & Reeves, C. A. (2000). Teacher responses to rewards and sanctions: Effects of and reactions to Kentucky's high-stakes accountability program. In B. Whitford & K. Jones (Eds.), *Accountability, assessment, and teacher commitment: Lessons from Kentucky's reform efforts.* Albany, NY: SUNY Press.

Katz, M. B. (1971). *Class, bureaucracy, and schools: The illusion of educational change in America.* New York: Praeger.

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record, 91,* 509-536.

Madaus, G. (1998). The distortion of teaching and testing: High-stakes testing and instruction, *Peabody Journal of Education, 65,* 29-46.

Mathison, S. (1991). Implementing curricular change through state-mandated testing: Ethical issues. *Journal of Curriculum and Supervision, 6,* 201-212.

Mathison, S., Ross, E. W. & Vinson, K. D. (2001). Defining the social studies curriculum: The influence of and resistance to curriculum standards and testing in social studies. In E. W. Ross (Ed.), *The social studies curriculum: Purposes, problems, and possibilities.* Albany, NY: SUNY Press.

McLaughlin, M. W. & Talbert, J. E. (2001). *Professional communities and the work of high school teaching.* Chicago:University of Chicago Press.

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing.* New York: Routledge.

Noble, A. J., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform: "Build it and they will come." *Educational Policy, 8*(2), 111-136.

O'Day, J. A. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review, 72*(3).

Ross, E. W. (2001). Diverting democracy: The curriculum standards movement and social studies education. In D. W. Hursh & E. W. Ross (Eds.), *Democratic social education: Social studies for social change.* New York: Falmer Press.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 15*(2), 4-14.

Strike, K. A. (1993). Professionalism, democracy, and discursive communities: Normative reflections on restructuring. *American Educational Research Journal, 30*(2), 255-275.

Thornton, S. (1991). Teacher as curricular-instructional gatekeeper in social studies. In R. Shavelson (Ed.), *Handbook of research on social studies teaching and learning*. New York: Macmillan.

Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal, 28* (3), 521-42.

Vinson, K. D. (1999). National curriculum standards and social studies education: Dewey, Freire, Foucault, and the construction of a radical critique. *Theory and Research in Social Education, 27*(3), 296-328.

Vinson, K. D., Gibson, R., & Ross, E. W. (2001). High-stakes testing and standardization: The threat to authenticity. *Monographs of the John Dewey Project on Progressive Education, 3*(2).

About the Authors

Sandra Mathison
College of Education and Human Development
University of Louisville
Louisville KY 40292
smathison@louisville.edu

Sandra Mathison is Professor of Education at the University of Louisville. She teaches evaluation and qualitative research methods and her research focuses on democratic and fair evaluation practices in schools.

Melissa Freeman
School of Education
University at Albany, SUNY
1400 Washington Avenue
Albany NY 12222
freeman@sover.net

Melissa Freeman is project manager of an interpretive study of the impact of high stakes testing in upstate New York. Her interests include theoretical and methodological issues in interpretive research and evaluation, democratic practices in schools, and critical social theories.

Acknowledgment

This publication is based on research supported by the National Science Foundation (Grant # ESI-9911868). The findings and opinions expressed herein do not necessarily reflect the position or priorities of the sponsoring agency.

Notes

1. While there is ample debate about whether teaching is a profession or not, and whether it ought to be considered a profession (see Strike, 1993) there are strong arguments for labeling teaching a profession (Darling-Hammond, 1990; Little, 1990; McLaughlin & Talbert, 2001). We adopt the view that teaching is a profession because it requires specialized knowledge and skills, especially as manifest in Shulman's notion of pedagogical content knowledge (1987) and contemporary theories of child development. In addition, teachers just as all other professionals are concerned simultaneously with both means and ends.

2. We wish to thank Kate Abbott and Kristen Campbell-Wilcox, our research collaborators on this project.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State

University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State Univeristy

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)

OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

American Institutes for Resesarch–Brazil
(AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

PDF file of this article available:

This article has been retrieved 709 times since September 25, 2003

Volume 11 Number 35 **September 25, 2003** **ISSN 1068-2341**

Communities of Practice and the Mediation of Teachers' Responses to Standards-based Reform

Chrysan Gallucci
University of California, Santa Barbara

Citation: Gallucci, C. (2003, September 25). Communities of practice and the mediation of teachers' responses to standards-based reform. *Education Policy Analysis Archives, 11*(35). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n35/.

Abstract

This paper evaluates the usefulness of a sociocultural approach for analyzing teachers' responses to the professional learning demands of standards-based reform policies. A policy-oriented case study of the practice of six elementary teachers who worked in two high poverty schools in a demographically changing district in the state of Washington is summarized. Key findings of that study conclude that *communities of teaching practice* are sites for teacher learning and are mediators of teachers' responses to standards-based reform. Characteristics of the communities of practice, including their relative strength and openness (to learning), influence the degree to which teachers work out negotiated and thoughtful responses to policy demands. The present paper discusses the efficacy of Wenger's (1998) theory of learning for the study of policy to practice connections.

Over the past decade, as the standards-based reform movement has swept the United States, the focus of education policy has shifted to the work of classroom teachers (Elmore, 1996; Thompson & Zueli, 1999). (Note 1) Researchers note that the content standards commonly associated with the reforms constitute a demanding curriculum for teacher learning (Borko & Putnum, 1995; Hawley & Valli, 1999; Thompson & Zueli, 1999). However, even in states that have placed high stakes on the improvement of student learning outcomes, scholars report that the reforms are not producing significant or large-scale change in teaching practice (Spillane & Zueli, 1999; Elmore, 2000). One major flaw in the design of standards-based policies is the insufficient attention that has been paid to the teacher learning that is necessary for instructional change to occur (Elmore, 1999; Darling-Hammond & McLaughlin, 1999). Thompson and Zueli (1999) argue that the problem with the implementation of government-driven systemic reform is a misunderstanding on the part of policymakers about the kind of transformative learning required by teachers if the ambitious content standards are to be realized in practice.

The purpose of this paper is to evaluate the usefulness of a sociocultural approach to this problem. Using the construct, *communities of practice* (Lave & Wenger, 1991; Wenger, 1998), I describe how the characteristics of professional communities mediate teachers' responses to reform policies. I summarize the findings from a group of teacher case studies that were developed during the 1999-2000 school year (Gallucci, 2002). In arguing that the implementation "problem" inherent in standards-based reform is about professional learning, I draw upon sociocultural theories of learning to aid my analysis of the practice of teachers who work in two high poverty schools. In this analysis, I ask: how do communities of practice mediate what teachers learn in response to reform efforts and, consequently, what shifts or changes do they make in their instructional practice? The treatment of findings is intentionally brief in this paper and is provided as a reference point for the theoretical discussion that follows. For a full explication of the evidence base of the larger, multi-level policy study, the reader is referred to Gallucci, 2002.

The research reported here builds on an earlier set of classroom-based case studies that focused on teachers' responses to standards-based instructional policies (EEPA, 1990). Those studies demonstrated that teachers' responses to curricular reform are likely to be modest, even when the teachers themselves believe they are making major changes in practice (Cohen & Ball, 1990). That team of researchers also found that teachers' responses and predispositions toward policies vary across a broad spectrum that ranges from active openness to the demands of new policies, to active resistance to them (Cohen, 1990; Ball, 1990; Wilson,1990; Wiemers, 1990). The studies suggested that local contexts—broadly conceived to include local conditions, interacting local policies, and teachers' own knowledge and assumptions about teaching and learning—powerfully shape teachers' responses to systemic policies. And increasingly, studies consider critical contextual dimensions of practice, especially the effects of professional relationships, *even* when the research focuses on subject-specific teacher learning. (Grossman & Wineburg, 2000; Franke & Kazemi, 2001.

There is a history of research regarding the relationship between teacher learning, teacher collaboration, and school improvement (Little, 1982, 1990, 1997; Rosenholtz, 1989; Johnson, 1990; Hargreaves, 1994; Louis, Marks & Kruse, 1996; Marks & Louis, 1999). And while there is much agreement that collaborative cultures create beneficial conditions for teacher learning, the nature of professional cultures and their connection to teacher learning have long been viewed as problematic (Lortie, 1975; Little, 1982; 1990; Hargreaves, 1994). Recently, some researchers have considered the role of *professional community* on teachers' work. Variation among professional communities has been found to influence the ways that teachers think about their practice. For example, McLaughlin & Talbert (2001) found that differences in qualities and characteristics among high school departments (professional communities that were either innovative or traditional) accounted for differences in the ways that teachers conceived of instruction for their increasingly diverse student bodies. A relatively small number of policy researchers have also demonstrated that professional teaching communities play a role in mediating teachers' responses to policy (Spillane, 1999; Coburn, 2001). In an analysis of the ways in which teachers make collective sense of dynamic reading policies through conversations that take place in their formal and informal professional affiliations, Coburn (2001) suggests that this sensemaking process mediates the nature of individual instructional change.

I move beyond single subject matter analyses in this work and suggest that the general characteristics of their communities of practice make a difference in how elementary school teachers respond to reform policies across content areas. Knowing more about the ways that communities of practice influence teachers' work enriches our understanding of the relationship between education policy and classroom practice.

Using Sociocultural Learning Theory to Study Teacher Learning in Context

I connected two bodies of conceptual and theoretical work in the framing of this study. First, I adapted ideas about the *embedded contexts* of teachers' work for use in this policy implementation study (McLaughlin & Talbert, 2001). (Note 2) McLaughlin & Talbert (2001) identified a layer of context between the classroom and the school organization as "teacher community and culture" (p.144). For the purposes of this study, I defined entities such as grade-level teams of teachers, teaching partners, and other configurations of teachers who work together as *potential* communities of practice (Wenger, 1998). Communities of teaching practice were conceptualized as the locus of (a) engagement in the actions of teaching, (b) interpersonal relations, (c) shared knowledge, and (4) negotiation of meanings about the work (Wenger).

I also took into account in this study the array of social and organizational variables that have the potential to impact teacher action in relation to policy intent. They included, (a) the social conditions of students' lives, (b) school-level organizational features (such as scheduling, school design features, school culture), (c) features of the community including parent culture, (d) professional contexts, and (e) district, state, and national policy environments. The assumption was that teachers may respond differentially to a set of policies based on the social, organizational, or political conditions of their work.

Second, social theories of learning provide a theoretical basis for understanding how teachers perceive policy environments, assign meaning to them, and extract insights from them. Sociocultural learning theories create a bridge between models of embedded contexts and the study of individual teacher learning within a reform environment. In general, these theories assume that learning is a phenomenon that is situated in and mediated by sociohistorical features of the environment such as language or artifacts (Lave & Wenger, 1991; Wertsch, 1991; Wenger, 1998). I draw here on a practice-oriented social theory of learning because the focus of this study was on learning as it occurs in the context of teaching practice.

Lave and Wenger (1991), and later Lave (1996), situate learning in communities of practice. They describe learning as shifts of participation in changing communities of practice (Franke & Kazemi, 2001). They suggest that these shifts—or learning—involve both changes in action and transformations of identity. The assumption is that individuals use means such as language, material tools or symbols, and interaction with other people to mediate their actions. Communities of practice collectively produce and are a source of cultural tools (or mediating factors) that affect individual teacher learning.

The study of the relations among (a) teachers' learning in communities of practice, (b) the organizational and social contexts of their work, and (c) teachers' instructional change calls for a theory of learning that links local practice to global supports or constraints on that practice. Wenger (1998) provides a framework for the analysis of communities of practice and their relationship to external structures. He locates communities of practice as a mid-level unit of analysis. He states that they are neither sites of specific, narrowly defined activities and interactions nor broadly defined conceptual aggregates that are abstractly social or historical (Wenger, 1998) (refer to Table 1). (Note 3) Wenger's framework suggests, rather, that the analysis of teacher learning (learning situated within and mediated by communities of practice) falls between minute interactions and activities and the world in aggregate. This theory can elucidate the potential *connections* between teachers' practice and standards-based reform measures.

Research Methods

A multi-level case study design was employed for this study (LeCompte & Preissle, 1993; Miles & Huberman, 1994). The study was conducted in an embedded set of policy environments (e.g., Washington State, Pinehurst School District, and two high poverty schools in that district—Rice Elementary and Maple View Elementary). Three teachers from each of the two schools were selected as the case study participants (refer to Appendix A for further description of the state, district, and school contexts). While background interviews and document collection were developed for the state, district, and school contexts, the teachers and their classroom practice were the subjects of focus for the study. The selected teachers taught a range of grade levels at the schools (1st through 5th grade) and their teaching experience ranged from 3 to 11 years.

I observed the teachers in their classrooms and across a variety of school settings especially in meetings and in other interactions with their colleagues over the course of a school year. I interviewed each teacher three times during the year and collected a variety of relevant documents such as curricular materials, lesson plans, and examples of student work. Interviews were also conducted with school principals, teacher specialists, and relevant district administrators.

Descriptive case summaries were developed for the district, school, and the six teacher cases. The teacher cases were analyzed individually using the coding system described in Table 1. I also conducted cross-case analyses of the teacher cases to develop interpretive understandings that helped explain the teachers' responses to standards-based reform (LeCompte & Preissle, 1993). Table 2 presents a summary of the major cross-case themes in each analytic category. An extended discussion of the research methods can be found in Appendix A.

Table 1
Analytic Codes Used to Develop Case Summaries

Analytic Code	Definition (Adapted from Wenger, 1998)
1. Communities of Practice	Defined as grade level teams of teachers, teaching partners, or other configurations of teachers working together that are potential communities of teaching practice. Defined here as the *most local group of teachers* with whom the case study teacher works out the daily demands of her work. Characterized by Wenger (1998) as having the following indicators: • Sustained mutual relationships, either harmonious or conflictual. • Shared ways of doing things (together). • Rapid flow of information and propagation of innovations and ideas. • Absence of introductory preambles, as if conversations and interactions were merely the continuation of an ongoing process. • Substantial overlap in participants' descriptions of who belongs. • Knowing what others know, what they can do, and how they can contribute to an enterprise. • Mutually defining identities. • Ability to assess the appropriateness of actions and products. • Specific tools, representations, and other artifacts. • Local lore, shared stories, inside jokes, knowing laughter. • Jargon and shortcuts to communication as well as the ease of producing new ones.

	• Certain styles recognized as displaying membership. • Shared discourse reflecting a certain perspective on the world.
Analytic Codes Used to Develop Case Summaries.	
2. Engagement in teaching practice Sub codes: instructional style (e.g., pedagogies, classroom management, interactions with students); curriculum and assessment (e.g., reading/writing, math); planning and organization; ideas and attitudes about practice.	One of the ways that teachers participate and belong to communities of practice (learn within communities of practice). Involves doing joint tasks, developing relationships and a shared repertoire. Could be meeting; talking; having time and places to do so; having or giving help; developing and defining competence; devising solutions and meanings; having stories about practice; gossiping; remembering; developing discourses; maintaining continuity over time; and constructing a learning trajectory.
3. Opportunities for Imagination Sub codes: district opportunities (e.g., curricular standards, assessment practices, professional development); the school as an opportunity, teacher-initiated opportunities.	The materials or resources that enable teachers to adopt other perspectives outside of their own bounded practice. Involves orientation to images of what could be (e.g., classes, curricula, videos, models, etc.); reflection (e.g., retreats, time-off, conversations, breaks in rhythm, etc.); and explorations or trying new things out (e.g., trying out new curricula, using ideas from an inservice, visiting other classes).
4. Alignment of practice with policy Sub codes: curricular policies (e.g., reading, math), assessment policies, other relevant policies.	The process that produces the ability to act with respect to a broad and rich picture of the world, to do something in concert with others, to embrace a bigger idea as part of our identity. Alignment involves making shifts or changes in practice based on a new idea or set of ideas. Alignment includes convergence around a common vision, coordinating practice with new standards or methods, or enforcement of new policies or procedures (i.e., by external structures).

Table 2
Implementation as a Learning Problem: Cross-Case Themes

Category	Cross-case Theme
Engagement in Practice	
Common Tendencies among Teachers across Cases	• Teachers were using the district-mandated curricula. • There was evidence of new or progressive ideas seeping into these teachers' practice. • The teachers saw themselves as caregivers for their students.
Variation among Communities of Practice	• New teachers in weak communities of practice followed the curriculum closely. • Strong and open communities made decisions about what to discard from their current repertoire. • Some communities were relatively open to new ideas and some were set in their current ideas. • Strong and open communities looked for instructional solutions to help ameliorate the social conditions of their students' lives. • Strong and closed communities tended to "blame" the students and their families for the ways that the conditions of their lives interfered with their schooling.
Opportunities to Imagine New Ideas	
Common Tendencies among Teachers across Cases	• Curriculum and assessment policies represented an opportunity for new learning. • There was a disparate array of inservice courses that represented the professional development opportunities for these teachers. • The teachers learned from each other.
Variation among Communities of Practice	• Strong, closed communities of practice were suspicious of new materials and tended to reject them as a source of new learning. • Strong, open communities examined the adoptions in light of their own practices and, thus, used them as an opportunity for learning.

	• New teachers in weak communities relied on the new curricula in a way that begged questions about the richness of this means of learning. • Decisions about inservice opportunities tended to be made at the individual or school level. Communities of practice were overlooked as a source of collaborative or embedded learning. • School-level decisions created opportunities (or not) for teachers to work together although some communities of practice did not take advantage of the opportunities. • Strong communities were more apt to influence each other and open communities tended to have a positive influence on learning within the communities.
Alignment Between Practice and Policy	
Common tendencies across Teachers 1. Curriculum policies 2. Assessment policies	• Curriculum adoption policies at the district-level overrode a focus on content standards. • Veteran teachers negotiated alignment with district mandates. • Early career teachers followed the adopted curricula and seemed to appreciate the structure. • Some classroom practice broadened as a result of the assessment content. • State and district assessments were driving curriculum and instruction toward test-related content. • Teachers think they know better (that they should be teaching to the individual child and not to a test).
Variation among Communities of Practice	• Strong communities talked about curriculum together. • Strong communities worked out some form of negotiated alignment with the district mandates. • Strong, but closed communities were predisposed to reject the district curriculum choices if they were in conflict with their current ideas about practice. They tended to work out a compliant alignment. • Strong, but open communities reviewed and worked with the new materials before they integrated them into their practice in a negotiated and thoughtful way. • Weak communities were compliant and relied on the district-mandated curricula. • All teachers across all communities of practice were responding to the high-stakes assessment policies. These policies cut across the variation patterns among the communities of practice.

Teachers' Response to Reform Policies: A Summary of the Findings

The purpose of this article is to discuss the usefulness of communities of practice as a construct for analyzing teachers' responses to reform policies. Therefore, in briefly illustrating the key findings of the study, I focus on the teachers' most immediate community of practice—that group of teachers with whom they work out the most pressing demands of their daily work (Wenger, 1998). The communities of practice in which the teachers participated varied along two important dimensions. First, some of the teachers in the study worked in what I characterize as *strong* communities of practice, in which teachers worked together in designing instruction and had a strong influence on each other's practice. Other teachers worked in *weak* communities of practice, in which teachers had much less influence over each other's practice. (Note 4) Second, the stronger communities varied along a dimension that could be characterized as relative *openness* to new ideas versus *insularity*, or being closed to outside ideas. In other words, a community of practice can have a strong influence on the practice of a group of teachers and that community may be strong in its unwillingness to entertain new or reforming ideas. (Note 5) Table 3 illustrates these dimensions.

Table 3
Key Dimensions of Difference Among the Communities of Practice

	Openness	Insularity

Strong	Teachers work together to design curriculum, plan lessons, and assess student work. Teachers negotiate actively with new policies.	Teachers work together, making curricular decisions and sharing responsibilities. They are set in their ways and oppose new policies.
Weak	Individual teachers design curriculum. They accept new policies, but lack community with which to create strong responses.	Teachers work alone and teach using methods that are familiar to them. They respond superficially to new policies, or tend to ignore them.

Communities of teaching practice are conceptualized as the locus of (a) engagement in the actions of teaching, (b) interpersonal relations, (c) shared knowledge, and (d) negotiation of meanings about the work (Wenger, 1998). Wenger characterizes communities of practice as having sustained relations (either harmonious or conflictual), shared ways of doing things, agreement concerning who are members, and shared stories, inside jokes, and other forms of shared discourse. For several teachers in this study, some or most of these characteristics were identifiable traits of their communities of practice. In the following section, I introduce one of the teachers from Rice Elementary who participated in this study. I describe the ways in which her work as a teacher is embedded in a set of overlapping communities of practice.

Illustrating Communities of Teaching Practice

Teachers at Rice Elementary School are not as isolated in their work as teachers have traditionally been portrayed (Lortie, 1975). Their rooms are clustered in three different pods—grouped roughly by grade level. The adults in the clusters form various kinds of working partnerships and friendships with each other. In that way, they form communities of practice with some of the members of their cluster and/or other educators in the building.

Sandra partners closely with the reading specialist who works with grades three through five at Rice, and who has an office near Sandra's classroom. They plan reading and language arts instruction for Sandra's class, co-coordinate the Washington Reading Corps (WRC) volunteer tutoring program, and plan special instructional units that they co-teach during vacation periods. She and the other third grade teachers also plan together with the reading specialist to coordinate reading curriculum across third grade. *"We plan with the reading specialist and we have set our whole schedule for third grade that way. There are four of us and we get together to plan our homework. We all send home the same things so we have some continuity."* Sandra works especially closely with the third-grade teacher next door. Over the years, they have developed several curricular units (for example, a unit of Northwest Native American tribes; a rain forest unit). *"We plan for the basal reading in our classrooms. We've planned the whole year out for that. And then, of course, we plan for science. We also meet with the math specialist who comes into our classrooms."*

Sandra's community of practice extends to the other teachers in her cluster, although in a slightly different way. The whole cluster might be considered a different, but closely related, community of practice.

> In our cluster, we like going to dinner once a month and celebrating birthdays or whatever. We enjoy it. We like each other's company; we are supportive of each other. We do things together. And say nice things to each other. We might bring a student over to another classroom for a while. And not that we are long lasting friends, but just that we make a point of getting together with the families and the spouses and just letting down. To laugh. It just makes everything so much better.

While it is not true of all the clusters in the building, in Sandra's cluster, "We have a pact that if any of us are going to leave or if anyone has any tension, we have to tell the others."

There are other people who participate in the work of teaching in and around Sandra's classroom. Those include parent volunteers, the coordinators of the WRC program, the special education teacher, the ELL teacher, the media teacher, and various teaching assistants. Her students, the building principal, the rest of the staff at Rice are also people with whom Sandra moves in and out of mutual engagement and joint work. It might be said that Sandra's working life is situated in a *nested* set of communities of practice (e.g., the individual classroom community, the third-grade team, the cluster, and the larger school community). For the purposes of this paper (that is, to describe their teaching practice in light of standards-based reform policies), I refer to the teachers' communities of practice as those in which they work out the most pressing demands of their work. For Sandra, that is her grade-level team and the reading and math specialists.

In the following paragraphs, I provide examples of the kinds of variation that were found among the case study teachers and their communities of practice. This summary highlights the key findings of the study and sets the stage for the theoretical discussion to follow. [For a complete explication of the findings of this study refer to Gallucci (2002).] The section is organized according to four important assertions regarding the mediating effects of communities of practice on teachers' responses to standards-based reform.

Assertion 1: Qualitative features of teachers' communities of practice affect teachers' interpretations of standards-based reform policies.

The ways that teachers made sense of the policy environment could be seen as they engaged in the work of teaching. For example, *all* of the teachers in this study were using district-mandated curricula. However, teachers who worked in weaker communities of practice, which included the newest teachers, were following the new reading curriculum verbatim often teaching with the curriculum guides in their laps. Within the same schools, the teachers who were working in strong and open communities of practice were integrating the newly mandated curriculum into their practice in a thoughtfully negotiated form of alignment with the policy. For example, one third-grade teacher and her grade-level teammates had developed a series of genre studies that had formed the content of their third-grade reading curriculum for several years. During the first year of the district reading adoption, these teachers used the adopted materials, but had *"opted out of the series for [our regular] reading rotations."* The teachers made their decision after using the new series, testing it out, and later negotiating their use of it based on their own knowledge and ideas about reading instruction. *"The stories in the reading series skip around. There may be a 'tall tale' here or there and maybe a historical fiction in the same theme. They fit the theme, but it's not really studying the genre in a chunk. And I feel, we all feel, that our kids learn better (I think we all learn better) when we can identify what we are studying."*

Noting the high-poverty settings in which they worked, the teachers in this study described themselves as *caregivers* for their students. The multi-level nature of the study provided evidence that the teachers' identity as caregiver was affected by school-level variables. In the school that had a strong programmatic vision—Rice Elementary had a Title 1 inclusion model—teachers took collaborative responsibility for developing a strong instructional program to support student learning. They added many school-wide supplemental supports to their instructional program such as an Accelerated Reader program, a volunteer tutoring program, and the Title 1 inclusion model. Maple View lacked a cohesive building-level vision and there, individual teachers tended to focus on the social conditions of their students' lives, in some cases, complaining that *"it's hard to keep a positive attitude when you are dealing with these kinds of kids all the time and you never get parent support and you have kids coming in tardy everyday"* (the reader is referred to Appendix A for brief descriptions of both schools).

In regard to this finding, the effects of the communities of practice were also observable. The teacher in a strong, open community at Rice Elementary was focused on developing a variety of instructional programs to meet the needs of her diverse students. However, a teacher in a strong but insular community of practice at Maple View Elementary was focused on the characteristics of her students that she felt made them unable to perform certain tasks or understand particular curricula. One teacher changed curricula to meet student needs; the other teacher expected the students to change in order to understand the curriculum. Differences in their communities of practice affected the ways that the teachers made sense of the social conditions of their students' lives and the associated implications for their work. These variations also affected the ways the teachers took up opportunities for learning.

Assertion 2: Characteristics of teachers' communities of practice affect their engagement with opportunities for learning.

Viewed at the level of policy design, Washington State and the Pinehurst School District had placed strong sets of ideas about curricula, in the form of content standards and assessments, into the environments in which these teachers worked. From the perspective of district officials, all of the teachers in the district had received "training" on their new adoptions and on the new forms of assessment. From the vantage point of the teachers, much of their opportunity for new learning consisted of a disparate array of inservice courses.

School-level organization and professional community influenced how opportunities for learning were taken up by teachers within their communities of practice. By virtue of particular policies such as school design, teacher assignment, and scheduling, schools influenced things like membership in teacher communities and time for communities to meet. At Rice Elementary, the strong school mission affected the teachers' generally educative stance toward their students; they believed that their Title 1 inclusion model worked, and that was evidenced in the high scores the students received on the state performance-based assessments. However, in both schools, there were examples of communities of practice being overlooked as potential sites for collaborative, sustained, or embedded professional development. Even in cases in which teachers were strongly influencing each other's practice, there was no evidence that anyone outside that community had purposely sought to design learning opportunities with the potential strength of that arrangement in mind.

The characteristics of the stronger communities of practice mediated the ways that the opportunities for learning were taken up by the teachers in this study. Because the communities could be either open to new ideas or closed off and insular, they affected how the teachers interpreted new ideas. Predictably, for example, a teacher in a strong but insular community of practice was suspicious of the district's new adoptions, as well as their curricular frameworks. She tended to reject them as a source of new learning. She commented, *"I guess a lot of language in the Essential Learnings* [district content standards] *is in the state curriculum. It is very general and very broad. Instead of starting from the basics and working out* [like we believe you should for first graders], *they are starting out broad and coming back."* In contrast, a teacher working in a strong and open community of practice had spent the summer working with fellow teachers to organize the 3rd grade curriculum to address the district's Essential Learning outcomes.

Assertion 3: Characteristics of teachers' communities of practice influence the kinds of changes that teachers make in their instructional practice in response to reform policies.

The teachers at both schools expressed overwhelm regarding the newly adopted curricula. Their focus on becoming familiar with these new texts and materials overrode any potential focus on content standards. However, veteran teachers who were members of strong communities of practice talked about the curriculum adoptions and worked together to develop *negotiated* alignment with the district mandates. Those communities that were typically open to new ideas interacted with the new materials and engaged their own knowledge to make decisions about their use (such as the teachers who decided to use their own genre studies for reading instruction).

Teachers who worked in strong communities that were more insular tended to respond superficially to the district mandates. They used the new materials enough to appear that they were conforming with policy, but they were actually doing what they had always done. For example, the first grade teacher at Maple View Elementary and her community of practice did not like the district's adopted reading series, but they did use it. *"Usually, I will introduce the leveled reader to the whole class as part of my morning routine. I haven't been overly fond of this series. If you did everything the series requires, it would take all day. I usually send copies of the leveled readers home with the kids."* This group of teachers was given waivers nine years ago to use another program. *"It's separate from the series, but we've been allowed to continue to use the Write to Read program. I hope they don't take it away because it's such a good program."* They use the "Write to Read" program three days a week for reading and writing instruction. When asked, these teachers described themselves as using the district's adoption.

Newer teachers, who in this study worked in weaker professional communities, were also compliant with district mandates. They relied on the district's choices as the primary source of curriculum materials and ideas. Their experience is an example of the effects of weak community among teachers. One can assume that the newer teachers wanted curricular guidance. The fact that it came in the form of textbook guides represents a lost opportunity for collaborative or sustained learning and engagement.

Assertion 4: Some reform policies overpower the characteristics of the teachers' communities of practice.

Teachers perceived the new assessment policies as having particularly high stakes for their work and as demanding their immediate response (McNeil, 2000; Whitford & Jones, 2000). Some aspects of these teachers' practice were broadened by their attempts to align their teaching with what they perceived to be the requirements of the new tests. For example, I observed teachers asking students to explain their answers to mathematical problems and using the six traits writing process methods that were promoted by the district. However, all of the teachers commented that the assessment policies were driving their practice toward test-related content. Third- and fourth-grade teachers especially expressed dismay at the amount of time that they were spending on test preparation activities.

The fourth-grade teacher at Rice Elementary said that she was teaching only things that were required on the new state assessment. *"It has taken a lot of freedom from us to teach the kinds of lessons we would like to teach. I don't consider many lessons that absorb any time at all unless I see that there is a very clear connection* [to the new tests]*."* She claimed that *"clearly 80% of the day"* was in some way connected to test preparation. At Maple View, the fourth/fifth-grade teacher added, *"It's definitely narrowing the curriculum. It's all focused on three subjects* [reading, writing, math] *and I think it's focused on specific types of skill."*

These powerful assessment policies cut across the differences in the communities of practice. The perception of the tests' high stakes for both teachers and students seemed to stun the teachers and they did little to achieve a negotiated alignment with the assessment (or the looming school accountability) policies. All of the teachers expressed deep concern about these negative effects. One teacher commented, "*We were always taught in school not to teach to the assessment but have your assessments reflect what you've been teaching. Our curriculum had not been addressed in a long time. Now they are addressing it, but they are focusing it so that it teaches to the assessment."*

Summarizing, as I studied the practice of these six teachers, I observed them working together in professional groups that I have described as communities of practice. Their interaction and identification with these communities—whether weak, strong, open or closed—mediated the teachers' individual responses to policy. In the following section, I discuss the relevance of sociocultural learning theory and the construct of communities of practice for policy implementation research.

Policy Implementation as a Learning Problem: A Theoretical Discussion

I began this research with a set of *policy to practice* questions that queried the response of elementary school teachers to state and district-level standards-based reform. I framed the interaction between standards-based policies and teacher practice as a problem of learning. I questioned the ways in which elementary school teachers, especially those who worked in high-poverty settings, were either supported or constrained by the intensity of the policy environment. Sociocultural learning theory provided a useful lens through which to pursue these questions.

Usefulness of Wenger's Theory

Although sociocultural perspectives on learning formed the theoretical framework for this study, I did not set out to study professional communities or communities of practice among teachers. As I observed the six teachers in this study and questioned them about their work in the context of standards-based reform, I noticed, however, that their explanations were often framed in collaborative terms. They talked about the ways that "they"—that is, themselves and the teachers with whom they worked—were developing responses to the social conditions of their students' lives and to the very present instructional policies associated with standards-based reform. I theorized that these teacher work groups were communities of practice and that they were mediating individual teacher learning and response to policy. Given this orientation, I turned to the work of Wenger (1998) as a tool for the analysis of my data.

Wenger's (1998) theory of learning falls among a broad set of social and psychological traditions that aim to keep the *individual* in play with the *social*, and action in a dialectic with structure (Giddens, 1984; Wertsch, 1985; Bruner, 1990; Lave & Wenger, 1991; Rogoff, 1995; Cole, 1996). These traditions view action and structure as mutually constitutive of each other, and as existing in transactional relation to one another. These sociocultural views of human phenomena lend themselves well to the study of structural conditions such as policy and their connections with individual actions such as teaching.

As noted earlier, Wenger (1998) addresses sociocultural learning theory within an organizational context. He notes "learning is an issue of sustaining the interconnected communities of practice through which an organization knows what it knows and thus becomes effective and valuable as an organization" (p. 8). This move to apply sociocultural learning theory to organizational contexts provides a particularly salient and useful framework for the study of the connections between education policy and classroom practice because teachers' work takes place in complex social and organizational contexts (districts, schools, and their communities).

There are other sociocultural theories of learning (and related constructs) that might provide useful tools for understanding the phenomena that I studied. Each of these orientations suggests a particular way to understand the problem of teacher response to policy. In large part, the choice of theory comes down to the relevance of what is foregrounded by a particular theoretical perspective in the context of the questions of interest.

Why *This* Sociocultural Theory?

An example of another sociocultural orientation is activity theory and the associated concept of appropriation (Herrenkohl & Wertsch, 1999). (Note 6) Herrenkohl and Wertsch distinguish between the mastery of a cultural tool (learning of the skills involved), and the appropriation of that tool as one's own. They apply the distinction between *mastery* and *appropriation* to the study of how children learn critical thinking skills through elementary science lessons. One can make a connection here between teachers' compliant use of curriculum materials, for instance, and a more thoughtfully worked out *negotiated alignment* with those instructional tools. In the first case, one could assume some level of mastery of the materials; in the second case, the teachers have appropriated the material as their own, modifying their use of them within the context of their particular practice.

The concept of appropriation applied in this manner is quite useful in understanding the individual teacher's response to a particular instructional policy. This lens foregrounds the cognitive development of the individual teacher, but is not as well suited to analyzing the ways that teachers learn in professional communities. The activity setting itself is context for individual appropriation; however, the focus of *this* study was on social organization at a broader level, including the interaction between individual actions, communal interactions, and the larger social structure.

Rogoff (1995) proposed a somewhat different unit of analysis for the study of learning and development. She suggested that learning takes place within activity systems that include three interdependent planes of analysis: participatory appropriation, guided participation, and apprenticeship. She described appropriation as participation (by the individual) in social activities and the process of that participation as "the substance of cognitive development" (p. 151). She saw guided participation as the interpersonal involvement of individuals and their social partners in social activity. She connected participation in smaller social groups to the accomplishment of larger institutional or cultural goals through a process of apprenticeship. Through apprenticeship, a novice learner is guided into increasingly more expert involvement in a broader social activity. (Note 7) The development of the individual learner within the sociocultural activity is the primary unit of analysis.

Each of these theoretical frames provides a potentially useful lens for the study of teacher learning and teacher response to educational policy environments. The application of Wenger's theory to this analysis, however, foregrounds a midlevel unit of analysis (communities of practice). His analysis of learning as it occurs within communities of practice (through engagement in joint work, exposure to new ideas, and efforts to make shifts in practice) adds much needed contextual information to what has previously been understood about teachers and their response to a standards-based reform environment. This orientation broadens our attention from the cognitive development and knowledge of the individual teacher, such as a teacher's knowledge about mathematics, to include the characteristics of the most local community context mediating that learning. It adds clarification to the phenomenon of within-school variation among teachers' responses to policy.

Subject-matter contexts have often been sites for the study of individual teachers' response to policy, especially to content-specific standards (EEPA, 1990; Jennings, 1996; Spillane, 2000). Previous studies, such as "The Case of Mrs. Oublier," were focused on how the individual teacher had understood and realized—or failed to understand and realize—reform intent in her teaching (Cohen, 1990). That study did not systematically examine the contexts of Mrs. O's work or her relation to them. There was an underlying expectation for enculturation in that case description—Mrs. O was expected to take up the mathematics reform in a particular way. The explanation for her implied failures did not explore the texture of her response as part of a broader picture of her teaching practice in relation to the various mediating contexts for her work. However, elementary school teachers face content standards across multiple subject matters that accumulate into intense demands for new learning (Gallucci, 1998; Spillane, 2000). In order to further our understanding of teachers' responses to standards-based policies, I studied the work of teaching across these multiple contextual demands. The use here of a particular sociocultural framework afforded a means to: (a) simultaneously study the policy environment, the contexts of teachers' work, and teachers' efforts to make meaning of the multiple dimensions of their teaching practice; and (b) to foreground teacher learning within communities of practice.

Previous work has clarified the transactional relationship between policy and the complex conditions of teaching practice, as well as the embedded nature of teaching practice (that is, teaching as embedded in multiple layers of organizational, social, and political contexts). However, the nature of these transactional relations, including those between change processes operating at the individual and organizational levels is not well understood. Sociocultural learning theory addresses exactly these kinds of transactional processes, and Wenger's framework focuses specifically on the ways that changes in practice (learning) are mediated by organizational structure and process. The findings of the present study suggest that this theoretical orientation to the problem of policy implementation has considerable heuristic value for researchers, policymakers, and practitioners.

Communities of Teaching Practice: Implications for Research, Policy, and Practice

By design, this research focused on a limited number of teachers in order to develop a deeper contextual understanding of their response to a particular set of instructional policies. This, of course, raises important questions about generalizability. The promise of this approach for understanding problems of policy implementation is supported, however, by the results of other case studies and survey research. Recent findings demonstrate the effects of social and organizational factors on teachers' responses to dynamic educational policies and the changing social conditions of their students' lives (see McLaughlin & Talbert, 2001; McNeil, 2000; Whitford & Jones, 2000; Coburn, 2001, for example).

The findings reported here clarify the need for further research regarding the use of this sociocultural construct for understanding teachers' responses to policies. I document that communities of practice among teachers vary along two important dimensions: (1) the relative strength or weakness of the community and (2) the relative openness of the community to engagement with new ideas about instructional practice. I hypothesize that these differences among the communities of practice affect teachers' interpretations of standards-based reform policies, their engagement with learning opportunities, and, subsequently, the kinds of changes that they make in classroom practice. These findings would be strengthened by research that further probes individual teachers' participation in communities of practice as well as across multiple communities of practice documenting evidence of learning from one professional setting to another (such as from one community of practice to another, or from a community of practice to a classroom).

A primary concern for educators and policymakers is the strengthening of existing and latent communities of practice. These findings suggest that strong and open-minded communities of practice represent *learning communities*. Further research is necessary to understand the conditions that enhance and sustain these kinds of collaborative structures among teachers.

Strengthening Communities of Practice

Educational leaders need information about how to *recognize* communities of practice among teachers. This theoretical approach to the problem of reform implementation suggests that learning is occurring *in practice*—whether we recognize it to be or not. The ability to see communities of practice and how they serve to mediate teacher learning and teachers' responses to policies is a natural first step toward harnessing that energy in a direction that supports positive instructional change (that is, change that leads to the improvement of student learning outcomes).

I distinguish here between communities of practice and more formal and time-limited entities such as task forces or teams. Teams of teachers, or other educators, may exist to accomplish a particular, predefined task (such as reviewing a particular curriculum or developing a strategic plan); they may or may not become communities of practice—entities in which teachers negotiate the meaning of their everyday work through their learning and identification with a community of other teachers. Communities of practice—unlike informal networks that may also pass information among friends or co-workers—create, expand, and exchange knowledge about their practice, as well as develop individual capabilities (Wenger, McDermott, & Synder, 2002).

Among the biggest barriers to harnessing communities of practice are time and other institutional structures. Nonetheless, there is much that could be done at the district and school levels in terms of design and planning to ensure that communities of practice develop among teachers. Many school buildings have structures that encourage such arrangements, such as pods, clusters, and grade-level teams. Schools that acknowledge and encourage vision setting, ongoing professional learning, and collaboration among teachers will enhance the probability that communities of practice exist as strong sites of professional growth. Schools that are organized in ways that encourage these activities (for example, schedules that provide collaborative planning time, and activities that require collaboration among the members of communities of practice) enhance the probability that communities of practice will flourish. Without such organizational support and conscious planning, communities of practice may languish, depending on the volunteer or spare time efforts of particularly energetic teachers.

These recommendations echo earlier calls for school-level organizational features that support *and reward* various forms of teacher collaboration (Darling-Hammond, 1996; Little, 1999). Darling-Hammond identified a number of activities that local groups of teachers might usefully engage in such as shared curriculum development, setting high standards for student work, and collective assessments of student learning. As noted by Little and affirmed by the work of McLaughlin and Talbert (2001), strong teacher learning communities question and challenge curricula, pedagogies, and outcomes for students. The findings reported here add to earlier suggestions that policymakers prioritize and support collaborative professional structures and search for ways to focus the work of teaching communities on positive educational outcomes.

Boundary crossings are opportunities for new learning for communities of teaching practice (Wenger, 1998). Classroom teachers might travel to curriculum committees that include members from various school sites. Returning to their home-school community of practice, these travelers bring new ideas. Another example of boundary crossing is found in the practice of *intervisitation* in Community School District #2 in New York City (Elmore & Burney, 1997). In that school district, teachers travel to lab sites where mentor teachers model exemplary teaching practices. Likewise, professionals who work as on-site staff developers travel between communities of practice, such as district offices to school sites, and they also have the opportunity to infuse local communities of practice with new information about their work.

Teacher leaders working from within communities of practice might also be the travelers or boundary crossers who bring new knowledge back to their community of practice. Support and encouragement of 'teachers as reformers' or 'teachers as curriculum developers' takes on added importance when viewed in this light. Teachers who are empowered to participate in reform-minded activities might infuse their community of practice with such spirit and activity (especially if they are given support in terms of time and learning experiences). Practitioners working together toward reform goals have been described as "apprenticing [themselves] to one another" in their efforts (Dutro, Fisk, Koch, Roop, & Wixson, 2002). These kinds of practice-based efforts to respond to reform lead to a certain *accountability of practice* that is essential given the kinds of expectations that standards-based reforms have placed upon classroom teachers.

Dimensions of Learning and Policy Design

Policymakers face some distinct limitations on what they can accomplish regarding professional communities. Previous research suggests that policies cannot, for instance, mandate what or how teachers learn (McLaughlin, 1987). They cannot mandate that teachers all work together in strong communities of practice or that they develop openness to new ideas. They likely cannot expect that systems or people will change in particular ways because of policy demands. They can, however, provide incentives and support for teachers to work together in communities of practice. They can focus attention and resources such as money and time on activities that engage teacher knowledge and that infuse communities of local practice with new ideas for their work. A good example of this type of activity, supported by government resources, were the local and state-level committees that engaged many classroom teachers in developing content standards and curricular frameworks during the early years of standards-based reform efforts (Gallucci, 1998; Dutro, et al, 2002).

As policies are designed and as they are taken up in local settings, there are some dimensions of learning in practice that bear attention on the part of educators and policy designers (Wenger, 1998). First, learning is enhanced when meaning making is balanced between reified perspectives (such as those enacted through legislative policy) on the one hand and participation on the part of local practitioners, on the other hand. Of course, some ideas are realized in policy and some are not, and those contribute to what gets learned. But, the big ideas of policy exist in concert with participation in practice. Wenger suggests that it is through this duality (reification and participation) that the process of negotiating *meaning*, or learning, takes place.

If, for instance, policy is mandated (reified) at state or national levels of the system, leaving little opportunity for negotiation, then there may be little chance to develop relevant meaning through participation. Although actions at the local-level, or street-level, might change the original intent of such policies, those actions consist too often of localized reactions to policy demands (Weatherly & Lipsky, 1977). The same participation on the part of practitioners might have broad application to the field if local response was a supported and expected outcome of education policies. For example, state policy makers could "give permission" to local districts, schools, and teachers to use any of a variety of strategies to develop locally sensible responses to broad visions for reform (Dutro, et al, 2002).

The *direction* of social energy is best balanced, then, with the *generation* of social energy. Policies that are generative and that invite teachers to engage practical knowledge and to negotiate local meaning, give rise to ongoing learning. Using the example provided above, many state-level policy makers proposed that content standards be developed, but when they left the decisions about the content of those standards to the professional community, including classroom teachers, they balanced the direction of social energy with the generation of social energy (Gallucci, 1998; Dutro, et al, 2002).

It seems clear from these findings that the teachers in communities of practice that generated negotiated responses to standards-based reform policies were involved in making good local sense of those policies for their students. On the other hand, when the balance of power between policy design and teacher negotiability favored external sources of ideas (such as with the new state assessment policies), teachers felt compelled to be compliant with mandates that they disagreed with and even strong communities of practice were powerless to make a difference. The teachers were in danger of becoming disengaged from their own work. In these cases, communities of practice became an overlooked source of creative energy to produce positive learning outcomes for students.

Designs for learning require the *power* to influence the negotiation of meaning at the local level (Wenger, 1998). The process of identification gives meaning to our membership in communities, but it does not define the importance of those meanings within larger social configurations. That process involves having the ability and legitimacy to define whose, or which, meanings count. Local meanings, for instance, may be extremely, even intimately, important to members of communities of practice. But they may carry little or no power within larger professional contexts.

This tension came into play in this study when the power and meaning accorded to state-level student assessments carried high stakes, such as media attention and consequences for local schools. In that situation the local meanings that teachers held existed in tension with the importance of externally developed standards for practice. Teachers were caught in the dilemma of reshaping their identity around a new set of ideas or negotiating a local response to those ideas, or both. Here issues of power came into play; the ability of teachers working within communities of practice to negotiate meaning made a difference in terms of their response to standards-based reform. The privileging of a global set of ideas over local meaning led, in some cases for example, to a non-participatory or compliant response on the part of teachers. The ideal would be to balance perspectives and allow for negotiability of meaning at the local level.

This study provides evidence that communities of practice among elementary school teachers are sites for professional learning and negotiation with reform policies. The challenge for those concerned with the improvement of educational outcomes, especially for students who attend high poverty schools, is to develop further awareness of the effects of local professional communities on teachers' practice. Future research is needed to provide educators and policy makers with guidelines for recognizing and strengthening existing communities of practice and for designing organizational structures that support their development.

Notes

1. Standards-based, or systemic, reform was conceived as an attempt to achieve policy coherence by aligning three areas of education policy: (a) high curricular standards and aligned assessments of student

progress; (b) standards for teacher education, licensure, continued professional development, and evaluation; and (c) support for schools to structure the time and conditions for student and teacher learning (Smith & O'Day, 1991; Knapp, 1997). State governments have taken an unprecedented lead over the decade of the 1990s in establishing curricular frameworks, related statewide student assessments, and systems for holding schools and teachers accountable for raising student outcomes.

2. Traditionally, the idea underlying "embedded contexts" is that individuals act within a set of nested environments that give meaning to, provide resources for, and shape that action. See Bronfenbrenner, 1978 or, more recently, McLaughlin & Talbert, 2001.

3. Wenger (1998) argues, "our actions do not achieve their meanings in and of themselves, but rather in the context of a broader process of negotiation. By starting with practice as a context for the negotiation of meaning, I do not assume that activities carry their own meanings" (p.286). Therefore, discrete activities, or systems of activities, are not the unit of analysis here.

4. *Weak* professional communities may be a misuse of the construct, *communities of practice,* however I use the term (weak) here to distinguish the characteristics of the professional affiliations that I observed among the teachers that I studied. This is an example of an area that requires further research.

5. These findings map closely onto the findings that McLaughlin & Talbert (2001) report regarding high school teachers and professional communities. We have converged upon a similar set of ideas about the nature of communities of practice across teachers who work in both high school settings and elementary schools.

6. I am using the term *activity theory* rather loosely here to encompass a wide range of post-Vygotskian theorists who proposed activity as the appropriate psychological unit of analysis for the study of learning. Herrenkohl & Wertsch (1996) were specifically concerned here with "mediational means" or the use of cultural tools in human action.

7. Rogoff (1995) applied this theory to the study of Girl Scout cookie sales and the study of young children's cognitive development in learning how to participate in that activity.

References

Ball, D. (1990). Reflections and deflections of policy: The case of Carol Turner. *Educational evaluation and policy analysis, 12,* 327-345.

Borko, H., & Putnum, R. T. (1995). Expanding a teacher's knowledge base: A cognitive psychological perspective on professional development. In T. R. Guskey & M. Huberman (Eds.). *Professional development in education: New paradigms and practices* (pp. 35-36). New York: Teachers College Press.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist,* 513-531.

Bruner, J. (1990). *Acts of meaning.* Cambridge, MA: Harvard University Press.

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational evaluation and policy analysis, 23,* 145-170.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational evaluation and policy analysis, 12,* 327-345.

Cohen, D., & Ball, D. (1990). Relations between policy and practice: A commentary. *Educational Evaluation and Policy Analysis, 12,* 347-353.

Cole, M. (1996). *Cultural psychology: A once and future discipline.* Cambridge, MA: The Belknap Press of Harvard University Press.

Darling-Hammond, L., & McLaughlin, M. W. (1999). Investing in teaching as a learning profession: Policy problems and prospects. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 376-411). San Francisco: Jossey Bass.

Denzin, N. K. (1978). *Sociological methods: A source book* (2nd ed). New York: McGraw-Hill.

Dutro, E., Fisk, M.C., Koch, R., Roop, L.J., & Wixson, K. (2002). When state policies meet local district contexts: Standards-based professional development as a means to individual agency and collective ownership. *Teachers College Record, 104,* 787-811.

Educational Evaluation and Policy Analysis (EEPA) (1990). Entire thematic issue (devoted to the findings of the California study of elementary mathematics). Vol. 12, no. 3 (Fall).

Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard Educational Review, 66,*(1) 1-26.

Elmore, R. F. (April, 1999). *Learning in classrooms, schools, and school systems: The problem of large-scale educational improvement.* The Dewitt Wallace Lecture presented at the annual meeting of the American Educational Research Association, Montreal.

Elmore, R. F. (2000). *Building a new structure for school leadership.* Washington, DC: Albert Shanker Institute.

Elmore, R. F., & Burney, D. (1997). *School variation and systemic instructional improvement in Community School District #2, New York City.* Unpublished manuscript, Learning Research and Development Center, University of Pittsburgh.

Franke, M. L., & Kazemi, E. (2001). Teaching as learning within a community of practice: Characterizing generative growth. In T. Wood, B. Nelson, & J. Warfield (Eds.), *Beyond classical pedagogy in teaching elementary mathematics: The nature of facilitative teaching.* Mahwah, N.J.: L. Erlbaum Associates.

Gallucci, C. (Oct.-Nov.,1998). *Policy and professional growth: Teachers' reconstructing practice in a climate of state reform.* Paper presented at the annual convention of The University Council for Educational Administration, St. Louis, MO.

Gallucci, C. (April, 2000). *Framing the study of teachers' response to systemic reform in high-poverty elementary schools.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Gallucci, C. (2002). *Communities of practice and the mediation of teachers' response to standards-based reform.* Unpublished doctoral dissertation, University of Washington, Seattle, WA.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Berkeley: University of California Press.

Grossman, P., & Wineburg, S. (2000). *What makes teacher community different from a gathering of teachers?* (Document O-00-1). Seattle, Washington: University of Washington, Center for the Study of Teaching and Policy (CTP) and Center on English Learning and Achievement (CELA).

Hargreaves, A. (1994). *Changing teachers, changing time: Teachers' work and culture in the postmodern age.* New York: Teachers College Press.

Hawley, W. D. & Valli, L. (1999). The essentials of effective professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice.* San Francisco: Jossey-Bass.

Herrenkohl, L. R., & Wertsch, J. V. (1999). The use of cultural tools: Mastery and appropriation. In I. E. Sigel (Ed.), *the development of mental representation: Theories and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Jennings, N. E. (1996). *Interpreting policy in real classrooms: Case studies of state reform and teacher practice.* New York: Teachers College Press.

Johnson, S. M. (1990). *Teachers at work: Achieving success in our schools.* New York: Basic Books.

Knapp, M. S. (1997). Between systemic reform and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research, 67,* 227-266.

Lave, J. (1996). Teaching, as learning, in practice. *Mind, Culture, and Activity, 3,* 149-164.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, UK: Cambridge University Press.

LeCompte, M.D., & Preissle, J. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). San Diego, CA: Academic Press.

Little, J. W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal, 19,* 325-340.

Little, J. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record, 91,* 509-536.

Little, J. (1997). *Excellence in professional development and professional community.* Working paper. University of California, Berkeley.

Lortie, D. (1975). *Schoolteacher: A sociological study.* Chicago: University of Chicago Press.

Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal, 33,* 757-798.

Marks, H. M., & Louis, K. S. (1999). Teacher empowerment and the capacity for organizational learning. *Educational Administration Quarterly, 35,* 707-750.

McLaughlin, M.W. (1987). Learning from experience: Lessons from policy implementation. In

R. Odden (Ed.), *Education policy implementation.* New York: SUNY Press.

McLaughlin, M.W. & Talbert, J. (2001). *Professional communities and the work of high school teaching*. Chicago: University of Chicago Press.

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis. (2nd ed.)*. Thousand Oaks, CA: SAGE Publications.

Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In J. V. Wertsch, P. Del Rio, & A. Alvarez (Eds.), *Sociocultural studies of mind*. Cambridge: Cambridge University Press.

Rosenholtz, S. (1989). *Teachers' workplace: The social organization of schools*. New York: Longman.

Smith, M., & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-268). Bristol, PA: The Falmer Press.

Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies, 31*, 143-175.

Spillane, J. P. (2000). A fifth-grade teacher's reconstruction of mathematics and literacy teaching: Exploring interactions among identity, learning, and subject matter. *Elementary School Journal, 100*.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis, 21*, 1-28.

Spradley, J. P. (1979). *The ethnographic interview*. Fort Worth, TX: Harcourt Brace Jovanonich.

Thompson, C. L. & Zeuli, J. S. (1999). The frame and the tapestry: Standards-based reform and professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice*. San Francisco: Jossey-Bass.

Weatherly, R., & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special education reform. *Harvard Education Review, 47*, 171-197.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.

Wenger, E., McDermott, R., & Snyder, W.M. (2002). *Cultivating communities of practice*. Boston, MA: Harvard Business School Press.

Wertsch, J. V. (1985). *Vygotskyand the social formation of mind*. Cambridge, MA: Harvard University Press.

Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press.

Whitford, B. L., & Jones, K. (2000). *Accountability, assessment, and teacher commitment: Lessons from Kentucky's reform efforts*. Albany, NY: SUNY Press.

Wiemers, N. J. (1990). Transformation and accommodation: A case study of Joe Scott. *Educational evaluation and policy analysis, 12*, 327-345.

Wilson, S. M. (1990). A conflict of interests: The case of Mark Black. *Educational evaluation and policy analysis, 12*, 327-345.

Yin, R. K. (2002). *Case study research*. (3rd ed.). Newbury Park: Sage Publications.

About the Author

Chrysan Gallucci
University of California, Santa Barbara

Email: chrysan@education.ucsb.edu

Chrysan Gallucci is an Assistant Researcher in the Gevirtz Graduate School of Education at the University of California, Santa Barbara. Her research focuses on educational policy—specifically, the connections between K-12 and Teacher Education reform policies and professional learning; leadership and learning; and the application of practice-oriented, sociocultural learning theories to issues of professional learning in organizational contexts. She also works with SRI, International on a study of policy issues related to the teaching profession in the state of California. Her other professional experiences include classroom teaching (both general and special education) and work with K-12 schools and school districts on systems change and professional development projects.

Appendix A: Research Methods

I used qualitative case study methods in the design for this study (Yin, 2002). The phenomenon of interest was the way that the dimensions of the policy environments, the individual teachers, and other social and organizational factors interacted with one another. I studied the ways in which teachers responded to policy, but I did not study all aspects of teachers' work.

Sample Selection and Settings

The study was conducted in the state of Washington. As legislated in 1993, the state had several components of standards-based reform in place, including content standards and performance-based assessments; policies aligning accountability measures, teacher certification and teacher education with the curricular reforms were under consideration at the time of this study.

The investigation took place in Pinehurst School District (PSD), a mid-sized urban and semi-urban school district located along the main western corridor of the state. The district was located in one the fastest growing areas of the state and several of its schools had rapidly increasing numbers of economically disadvantaged or non-English speaking immigrant students. Pinehurst School District served approximately 21,500 students at the time of this study (an increase of about 33% over the 1990s). The district-wide free- and reduced-price lunch (FRL) population was about 10% in 1980 and was up to 40% in the 1999-2000 school year; it was over 50% in the elementary schools. Of the 21 elementary schools in the district, 13 were school-wide Title 1 eligible. [The Pinehurst community had suffered a severe economic downturn over the decade of the 90s due to the closing of several industrial plants.] Approximately 20% of the total students were receiving services for English Language Learners—the majority of those students were Spanish or Russian language dominant. In terms of ethnicity, the district had a stable pattern of about 85% White students. The percentage of African American students was stable at about 4%, but the Latino/a student population had risen from 2.5% to 7% over the decade of the 1990s.

Pinehurst had responded quickly and decisively to the state-level standards-based reform measures. Over a period of four years, PSD had centralized its curricular policies through three new content area adoptions (reading, science, and math). The mandated use of a specific reading series, for example, was considered controversial because prior to 1997 decisions about reading materials and related pedagogical practices were made at the school level. In addition, the district had developed an aggressive response to the new state-level student assessments, adding its own assessments and test-preparation requirements to those of the state. Washington State, in conjunction with the Riverside Publishing Company, developed the Washington Assessment of Student Learning (WASL). The test was in part a performance-based measure that was intended to test student outcomes on the legislature adopted curriculum standards. The test was administered in grades 4, 7, and 10; most districts, Pinehurst included, began administering the test in the spring of 1997 in English Language Arts and Mathematics.

Two schools that were characterized by high levels of poverty were selected for the study. I chose *high poverty* schools because the social conditions of students' lives and the generally low scores on standards-based assessments in such schools make them sites of particular interest for policy implementation research. These two schools were recommended by district personnel for the variance in their instructional approaches to school improvement, school organization, and professional cultures. The student populations of the schools were slightly more diverse than the district's overall student population, however, the students in both schools were predominately poor and White (69%-75%).

Maple View Elementary reported the following student data in 2000: 75% White, 15% Hispanic [district terminology], 5% African American, 3% Asian, and 1% American Indian students. At Maple View Elementary about *90%* of the students received free or reduced price lunch (1999-2000). The school had high mobility rates such that about *1/3rd of the students turned over in the first three months of school.* Less than 45% of the students at Maple View met or exceeded state standards on the new performance-based assessments in reading and mathematics (between 1998 and 2000 reading scores rose from 36% to 45%; math scores rose from 19% to 31%). Program delivery services at Maple View were organized in relatively traditional models (for example, students who received special services of any kind were pulled out of the general education classroom and moved to other locations in the building).

Rice Elementary reported in 2000 that 69% of its students were White, 14% Hispanic, 10% African American, 4% Asian, and 3% American Indian. At Rice Elementary about *70%* of the students received free- or reduced-price lunch and the school had a *mobility rate of about 40% per school year (1999-2000).* Test scores at Rice Elementary had climbed significantly over a three-year period (for example reading scores had climbed from 29% to 78% of the students at or above the state standards; math scores also increased but not as dramatically). This phenomenon was attributed largely to a redesigned Title 1 delivery model that brought reading and math specialists into regular classrooms.

Three teachers each from Maple View and Rice elementary schools agreed to participate in the teacher case studies. I purposely sampled teachers who were early in their career (2-4 years) and those who were experienced teachers (more than 7 years teaching) in order to compare teacher perceptions and experiences across a range of early to late teacher careers. In each school, I selected at least one teacher at the 3rd grade and 4th grade levels (these were the most highly tested grade levels). I selected teachers at both primary and intermediate grades at each school in order to balance my findings across grade levels in the schools. I also talked with the principals about my goals for the study (e.g., the study of teacher learning and standards-based reform; the need for teachers willing and able to articulately describe their work) and checked my selections with them before making my final decisions.

Data Collection

This inquiry was conducted using policy-oriented case study and ethnographic field methods. Data collection methods focused on both the policy environments and classroom practice, with emphasis placed on district, school, and teacher levels of the policy system. I analyzed state documents related to K-12 standards-based reform in order to provide state level context for the study. I relied on the Center for the Study of Teaching and Policy (CTP) case study development in Washington State as a source of additional information (I was associated with CTP during the time of the study).

I conducted semi-structured interviews with seven district personnel and I interviewed each school principal two times during the study. At the district level, I interviewed the Superintendent, two assistant Superintendents (Curriculum and Instruction and Learning Support), the Director of Human Resources who was also responsible for professional development, and three curriculum specialists. At each school, I interviewed school-based specialists such as the reading specialists and special education teachers. In addition, I collected and reviewed a variety of district documents, videotaped professional development materials, and demographic data as well as school mission statements, Student Learning Improvement Plans, school-level student outcome data and other materials appropriate for the goals of the study.

In developing the teacher case studies, I utilized ethnographic field methods including the collection of in-depth field notes and multiple teacher interviews over time (Spradley, 1979; LeCompte & Preissle, 1993). I observed each of the six case study teachers as they taught their classes on a minimum of six different days during the school year (1999-2000). These observations ranged in time from 2 hours to 6.5 hours with the average observation lasting one half of a school day. I also sat in on teacher meetings, individual teacher planning sessions, informal conversations, and lunchtime activities. I interviewed each of the teachers three times over a period that extended from January of 2000 through June of 2000 using semi-structured, in-depth interview protocols (Spradley, 1979). The interviews were typically one hour in length. All interviews were tape recorded and professionally transcribed in verbatim text for later analysis. I collected curricular materials, teacher developed lesson plans, and examples of student work. This use of multiple methods of data collection was one form of triangulation, ensuring that multiple data sources would balance findings and protect against reliance on a particular source (Denzin, 1978).

Data Analysis

Following ethnographic and interpretive traditions, data analysis for this project was ongoing and iterative. Formal steps in the data analysis process began with (1) a re-reading all of the raw data and (2) jotting notes and observations in the margins of the interview transcripts, field notes, and documents. My notes were based on the theoretical framework with which the study was initiated and the constructs described by the participants of the study (LeCompte & Preissle, 1993). I used this early scan of the raw data to build an inductive understanding of what was in the data.

I developed analytic case summaries for each of the six teacher cases using the coding system that I had developed to organize my findings (refer to Table 1 for the analytic codes). I also developed descriptive case accounts of both schools and the school district. Finally, I analyzed the data across the six teacher case accounts and within each major analytic category to develop interpretive understandings that explained the responses of the teachers to standards-based reform (LeCompte & Preissle, 1993). Table 2 presents a summary of the major cross-case themes in each analytic category.

To ensure that my interpretations of the data matched the reality of the participants in the study I incorporated the following procedures into the research process. First, during data collection, I provided the teachers with photocopies of the field notes that I collected in their classrooms and asked them if the notes were accurate. Second, during the data analysis process, I provided the teachers with their own case summaries and asked for feedback regarding the accuracy of my descriptions and interpretations. One of the teachers met with me to discuss her case account and two teachers sent email feedback regarding their case summaries. I followed up with the remaining three teachers and they confirmed that the case summaries were representative of their teaching practice.

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

David C. Berliner
Arizona State University

Greg Camilli
Rutgers University

Linda Darling-Hammond
Stanford University

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Gustavo E. Fischman
Arizona State Univeristy

Richard Garlikov
Birmingham, Alabama

Thomas F. Green
Syracuse University

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Ontario Institute of Technology

Patricia Fey Jarvis
Seattle, Washington

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Les McLean
University of Toronto

Heinrich Mintrop
University of California, Los Angeles

Michele Moses
Arizona State University

Gary Orfield
Harvard University

Anthony G. Rud Jr.
Purdue University

Jay Paredes Scribner
University of Missouri

Michael Scriven
University of Auckland

Lorrie A. Shepard
University of Colorado, Boulder

Robert E. Stake
University of Illinois—UC

Kevin Welner
University of Colorado, Boulder

Terrence G. Wiley
Arizona State University

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Maria Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil
(AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

PDF file of this article available:

This article has been retrieved 838 times since October 7, 2003

Volume 11 Number 36 **October 7, 2003** **ISSN 1068-2341**

English Learners in California Schools: Unequal resources, unequal outcomes

Patricia Gándara
University of California, Davis

Russell Rumberger
University of California, Santa Barbara

Julie Maxwell-Jolly
University of California, Davis

Rebecca Callahan
University of California, Davis

Citation: Gándara,P., Rumberger, R., Maxwell-Jolly, J. and Callahan, R., (2003, October 7). English Learners in California Schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives, 11*(36).Retrieved [Date] from http://epaa.asu.edu/epaa/v11n36/.

Abstract

The *Williams vs the State of California* class action suit on behalf of poor children in that state argues that California provides a fundamentally inequitable education to students based on wealth and language status. This article, an earlier version of which was prepared as background to that case, reviews the conditions of schooling for English learners in the state with the largest population of such students, totaling nearly 1.6 million in 2003, and comprising about 40 percent of nation's English learners. We argue, with evidence, that there are seven aspects of the schooling of English language learners where students receive an education that is demonstrably inferior to that of English speakers. For example, these students are assigned to less qualified teachers, are provided with inferior curriculum and less time to cover it, are housed in inferior facilities where they are often segregated from English speaking peers, and are assessed by invalid instruments that provide little, if any, information about their actual achievement. We end with suggestions for ways in which teachers, administrators, and policymakers can begin to address these inequities, even while legal remedies may remain in the distant future.

Introduction

The fact that the United States remains an immigrant nation is nowhere more apparent than in our public schools where an increasing percentage of students are English learners. In 2000-01 these students represented ten percent of all students in the United States, and 25 percent of California's public school population (Kindler, 2002). In the nation they numbered 4.6 million; while in California alone, they were more than 1.5 million students. Most English learners both in the U. S. (79%) and in California (83%) speak Spanish as their primary language. The second largest language group in both California and the U.S. is Vietnamese, however they account for only 2 percent and 2.5 percent respectively. California is also by far home to more limited English students than any other state. Thirty-three percent of all of the nation's English learner students live in that state; the next largest concentration is in Texas, with 12 percent of the total (Note 1). How best to educate these students continues to be a highly controversial topic and the source of considerable policy debate. However, with such a large population of English learners, it is surprising how little attention is actually paid to the basic learning resources these student receive in California, and in the nation.

English learners are distributed throughout the schools, from kindergarten to grade 12. One out of four students in the public schools in California is an English learner, but one out of three of the students in the elementary grades lacks proficiency in English (Rumberger & Gándara, 2000, Table 1). There are very few California schools that report having no English learners among their student population (see Table 14). Today, the typical California school is composed of both English learners and English speakers, and in many schools more than one-quarter of the student body is not fluent in English. Although most English learners are found at the elementary school level, a larger proportion of English learners (hereafter also referred to as ELs or EL students) is found in secondary schools than commonly believed. More than 18 percent of California's secondary school students (500,000 plus) are English learners (Rumberger & Gándara, 2000, Table 1). Proportionately, the number of English learners in secondary schools has been growing at a faster rate than the number in elementary schools (California Department of Education, Language Census 2001). The increase in the population of these secondary level English learners presents a particular challenge for both the students and the schools that serve them. This is principally because older children have less time to acquire both English and academic skills in order to get ready for high school graduation and to prepare for post-secondary options. Unfortunately the unique needs of these older EL students are often even more overlooked than those of their younger peers. This article assesses the condition of education for English learners in California, and, we believe, has significant implications for the nation.

Organization of this Article

We begin this article with an examination of the achievement data on English learners in California. We think it is first important to establish the degree to which these students' achievement represents a challenge to the overall productivity and welfare of the state's education system. In other words, we attempt to make the case that the achievement gaps are so wide that they threaten the well-being of the state and its economy, and therefore should be a concern to everyone. We then follow with a discussion of seven factors that we argue contribute significantly to this situation. These seven factors are not exhaustive of the problems faced by EL students, nor can they be neatly compartmentalized. Some, like the shortage of skilled teachers, represent both *input* shortcomings (e.g., insufficient numbers of qualified teachers) as well as *process* problems (e.g., inadequate instruction in the classroom) simultaneously. Therefore, we present these factors roughly in the order in which we think they affect the condition of schooling for English learners and are amenable to policy intervention. We end with a set of recommendations for addressing these issues. While this work grew out of a major class action lawsuit in California, we acknowledge that legal remedies almost certainly lie in the fairly distant future, and more general social change perhaps beyond that. Therefore our recommendations speak directly to the kinds of actions that school personnel might undertake in the shorter term.

Achievement of English learners

Data from a variety of sources reveal that the academic achievement of English learners lags considerably behind the achievement of English background students. We examined the achievement of English learners using a number of different measures and data sets –including data from the Early Childhood Longitudinal Study (ECLS) of the U.S. Department of Education, the American Institutes for Research Implementation of Proposition 227 Study (Parrish, et al., 2001; 2002) and the California Department of Education published data. (Note 2) At the same time that we present analyses of existing data on student achievement for English learners, we do so fully acknowledging the serious limitations of achievement scores based on tests administered in English to students who do not speak English well, or at all. We discuss below the issues associated with such assessment.

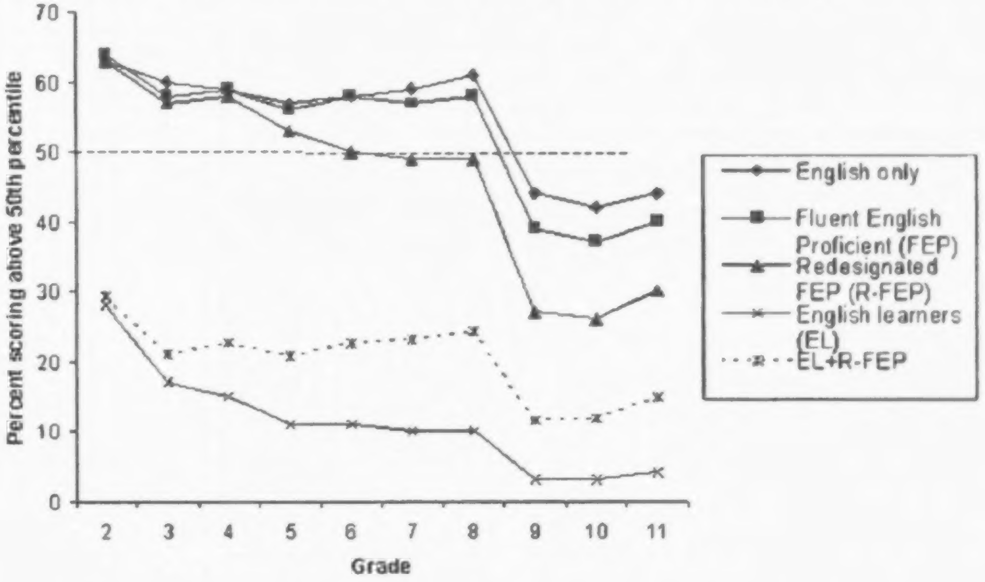
Stanford 9 Achievement Scores

Between the years 1998 and 2002, the state used the Stanford Achievement Test, Version 9 (SAT9)—a national, norm- referenced, English-only achievement test—as the primary means to assess the academic achievement of California's students. (Note 3) In spite of the fact that we disagree with the state's decision to use this test for students who do not speak enough English to understand it, we provide an analysis of the achievement of English learners vis-à-vis their English-speaking peers as these same test scores are routinely reported as accountability measures in the state.

A persistent gap in test scores is a major factor in the school experience of English learners. As a group they continue to perform more poorly than English-speaking students throughout their entire school career. This is clearly illustrated by the SAT 9 English reading scores across grade levels (see Figure 1). As expected, English learners who, by definition, are not yet proficient in English, have low reading scores across all grade levels. Language minority students who enter school already proficient in English (Fluent English Proficient or FEP) start out comparable to native English speakers, but by third grade they fall behind and never catch up. Students who enter the schools as English learners and who are subsequently reclassified as proficient (R-FEP), also start out comparable, but by 5th grade they fall below native English speakers, and by 7th grade they fall even further behind these students. Such results challenge the belief that if English learners simply demonstrated "proficiency" in English –as defined by early scores on the SAT 9 test—the achievement gap would disappear.

Even though this analysis shows a sizeable and growing achievement gap between English speaking and non-English speaking students across grade levels, there are some suggestions in the data that the gap has narrowed slightly in recent years. To investigate this issue, we examined SAT9 reading test scale scores between the years 1998 and 2001 compiled by Parrish et al. (2002) as part of the American Institutes for Research (AIR) year 2 evaluation of proposition 227. Scale scores show growth in achievement over time based on a common metric. Thus it provides a good indication of the amount of learning that has taken place over time.

Figure 1. 2001 California SAT9 Reading Test Scores by Grade Level and Language Background



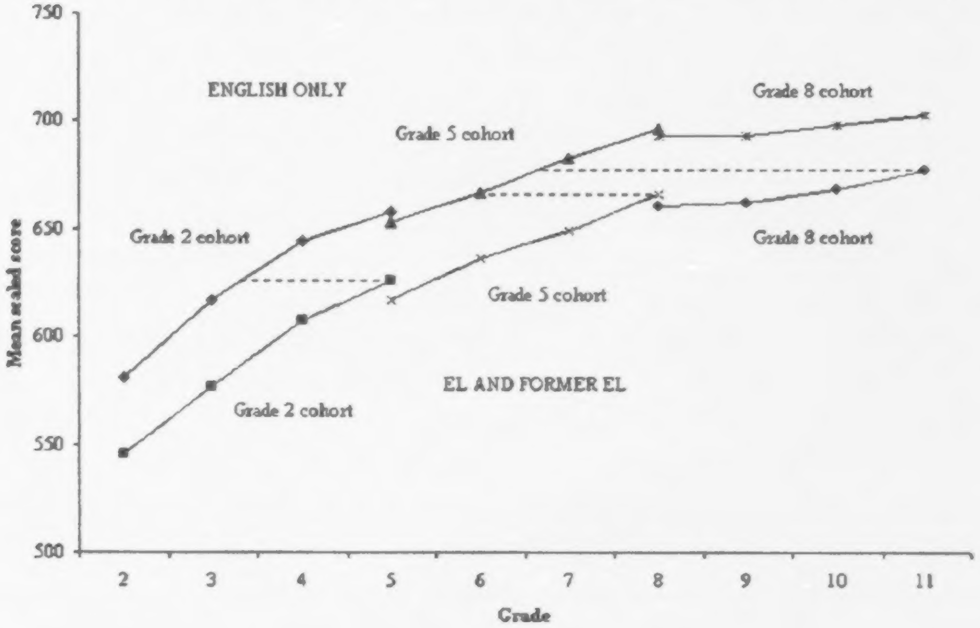
SOURCE: California State Department of Education, *California Standardized Testing and Reporting (STAR) Program*. Retrieved February 7, 2002 from the World Wide Web: http://star.cde.ca.gov/star2001/default.htm

The AIR evaluation team had access to individual student test scores for all the students in California for the years 1998 through 2001 by language classification. (Note 4) The evaluation team examined changes in test scores between 1998 and 2001 for each grade level and for three synthetic cohorts of students: (Note 5) (1) a cohort of students who were enrolled in grade 2 in 1998, grade 3 in 1999, grade 4 in 2000, and grade 5 in 2001; (2) a cohort of students who were enrolled in grade 4 in 1998, grade 5 in 1999, and grade 6 in 2000, and grade 7 in 2001; and (3) a cohort of students who were enrolled in grade 8 in 1998, grade 9 in 1999, grade 10 in 2000, and grade 11 in 2001. In order to compare non-overlapping cohorts, we replaced the second cohort with one that began when students were enrolled in grade 5 in 1998. One of the innovations of Parrish and his colleagues is that they compared English-only students with the combined group of current English learners and former English learners who were reclassified as Fluent English Proficient (R-FEP) in order to better assess the progress of all students who first entered California schools as English learners. Because an increasing number of EL students become proficient in English as they progress through school and are reclassified as fluent English speakers, the number of EL students tends to decrease among older grade cohorts while the number of R-FEP students tends to increase.

The results, illustrated in Figure 2, again show a sizeable achievement gap between English only students and current/former English learners. Both groups show more achievement growth in the early years than in the later years, which reflects the increasing difficulty of learning higher levels of more academic English

(Scarcella & Rumberger, 2000). The data show a slight narrowing of the achievement gap across all three cohorts, as Parrish, et al. note in their evaluation study (Parrish, et al., 2002, page III-15). For example, the achievement level of English only students improved from 581 points in grade 2 to 658 points in grade 5, an increase of 77 points, while the achievement level of English learners and former English learners improved 80 points. As a result, the achievement gap narrowed by 3 points. Among all three cohorts and three subjects (reading, language, and math), the 227 evaluation team found that the achievement gap narrowed by 1 to 8 points (Parrish, et al., 2002, Exhibits 10, 13, 16).

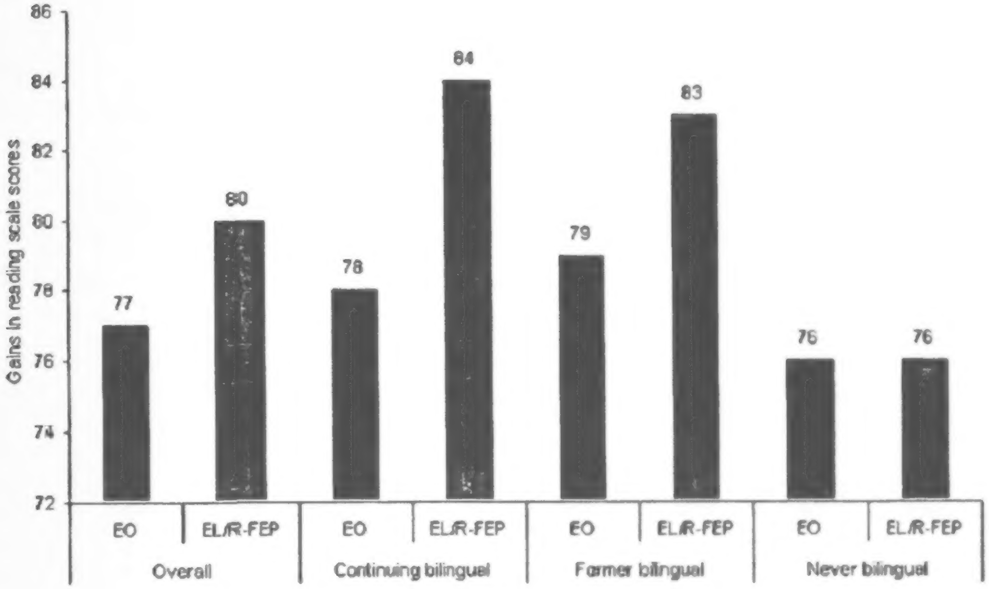
Figure 2. SAT 9 Reading Scores by Grade Cohort and Language Classification, 1998-2001



SOURCE: Parrish, et al. (2002), Exhibit 1.

It is interesting to note that the greatest achievement growth for the grade 2 cohorts occurred in schools that offered bilingual instruction before Proposition 227 or continued to offer bilingual instruction after Proposition 227 (Figure 3). In addition, the slight narrowing of the achievement gap between English only and EL and former EL students noted above was due to reductions in the achievement gap in those two types of schools, while in schools that never offered bilingual education, there was no reduction in the achievement gap.

Figure 3. Reading Achievement Gains for Grade 2-5 Cohort by Language Group and Instructional Model



NOTE: EO is English only; EL is English learner; R-REP is reclassified Fluent English Proficient (formerly EL).
SOURCE: Parrish, et al. (2002), Exhibits 10 and 19.

Despite these improvements, the achievement gap remains large and increases at the higher grades. To illustrate, in grade 5, when many students have completed elementary school, the left-most horizontal line in Figure 2 shows that current and former English learners are reading at the same level as English only students between grades 3 and 4, a gap of about one and one half years. By grade 8, when most students have completed middle school, the next horizontal line shows that current and former English learners are reading at the same level as English only students in grade 6, a gap of about 2 years. By grade 11, the right-most horizontal line shows that current and former English learners are reading at the same level as English only students between grades 6 and 7, a gap of about 4 and one half years.

California High School Exit Exam

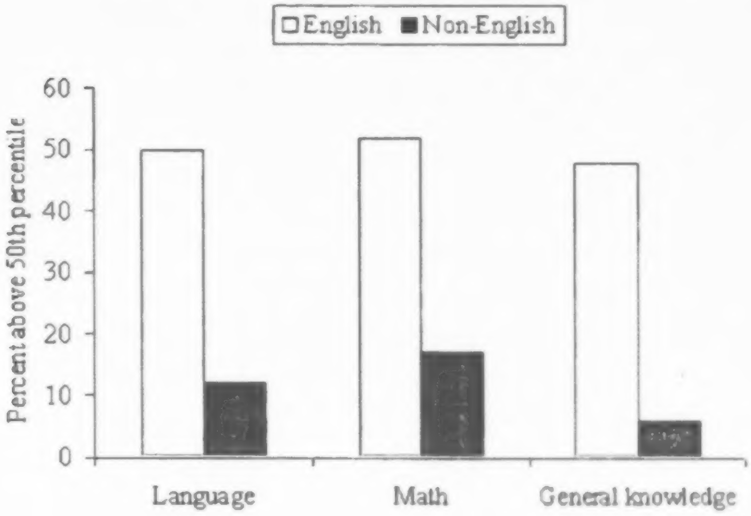
The California High School Exit Exam (CAHSEE) is a major element of California's education accountability system. All students in the class of 2004 and beyond were expected to pass the exam in order to receive a high school diploma until very recently when the State Board of Education authorized a delay in implementation of sanctions until 2007. The data provide a hint as to why the class of 2004 received this 11th hour reprieve. The exam is a standards- based, criterion-referenced test that is designed to ensure that all California high school graduates have a similar set of fundamental skills in English language arts and mathematics (California Education Code section 60850-60859). The need for improving the education provided by California's high schools is undeniable. Although accountability measures may be necessary to this effort, there is early evidence that the CAHSEE presents exceptionally high stakes for EL students.

Although the test is a basic skills examination pegged to early high school standards, (Note 6) by the end of their sophomore year, only 48 percent of students from the class of 2004 had passed it. However, only 19 percent of English learners had passed at this same point (California Department of Education, 2002, Attachment 1).

School Readiness

One reason for the underachievement of English learners is that they begin school significantly behind their English-speaking peers. Data from the Early Childhood Longitudinal Study (ECLS) show that about half of California kindergartners from English speaking backgrounds scored above the 50th percentile in fall assessments of language, mathematics, and general knowledge. However, no more than 17 percent of kindergartners from non-English speaking backgrounds scored above the 50th percentile (see Figure 4). One reason for this disparity is that many English learners begin school without a sufficient understanding of oral English that English background students acquire naturally in their home environment. According to the ECLS data, more than 60 percent of English learners who entered California kindergartens in the fall of 1998 did not understand English well enough to be assessed in English. And even after one year of school, 38 percent of the students were still not proficient enough in English to be assessed. (Note 7)

Figure 4. Cognitive Skills of California Beginning Kindergartners by Language Background, Fall 1998



Note: Results are weighted (C1CW0).
SOURCE: ECLS base year data for California public school kindergarteners (N=2826).

Teachers and schools make judgments about students' abilities based on the information available to them, including test scores. Schools make class placement decisions based, at least in part, on students' standardized test scores. Moreover, when the teacher does not speak the language of the child, cannot communicate with the child's family, and has little other information to rely on, test scores can take on even greater importance. Students who score low on tests are likely to be placed in remedial education, even though such a placement is unlikely to help students close the educational gap with their mainstream peers. (Gottlieb, Alter, Gottlieb, & Wishner,1994; Skirtic, 1991). In *Hobson v Hansen* (269 F. Supp.401, 490; DDC 1967), the Washington DC Superior Court noted in a major test case on the viability of curriculum tracking as an educational practice that "a sixth grade student nourished on a third-grade curriculum is apt to finish the year with a third-grade education. . .".

Conditions of Inequity for English Learners

The achievement gap between English learners and their English- only counterparts can be attributed, in part, to a number of inequitable conditions that affect their opportunities to learn. Our own research, combined with a review of the research of our colleagues, leads us to identify seven primary areas in which these students appear to receive a significantly inferior educational experience, even when compared to other low-income students in the public schools.

(1) Inequitable access to appropriately trained teachers

English learners are more likely than **any other** children to be taught by teachers with an emergency credential. There is reason for concern about the low percentage of teachers who are qualified to teach these students. An increasingly large body of research has established that teachers with good professional preparation make a difference in students' learning (Darling- Hammond, 2002; Haycock, 1998; Sanders & Horn, 1995; Sanders & Rivers, 1996). Moreover, a recent study conducted in Los Angeles Unified School District (LAUSD) investigated the relationship between English learner student achievement gains and the credential held by the teachers who taught them in 29 schools and 177 classrooms with large numbers of EL students. Hayes and Salazar (2001) found that "state/district authorization of teachers does have an impact on student outcome. For example, [Model B (Note 8)] students of teachers holding no state or district authorization achieved largely negative or very small positive. . . . adjusted gains in reading and language" (pp. 37-38). (See Table 1). A follow up study of grades 1 – 3 classrooms in the same schools during the subsequent school year (2001) found again that "students of credentialed teachers out-performed students of emergency permitted teachers" (Hayes, Salazar & Vukovic, 2002, p. 90).

Table 1. Actual and Adjusted Gains by Teacher Authorization Grade 2, Selected Schools, LAUSD

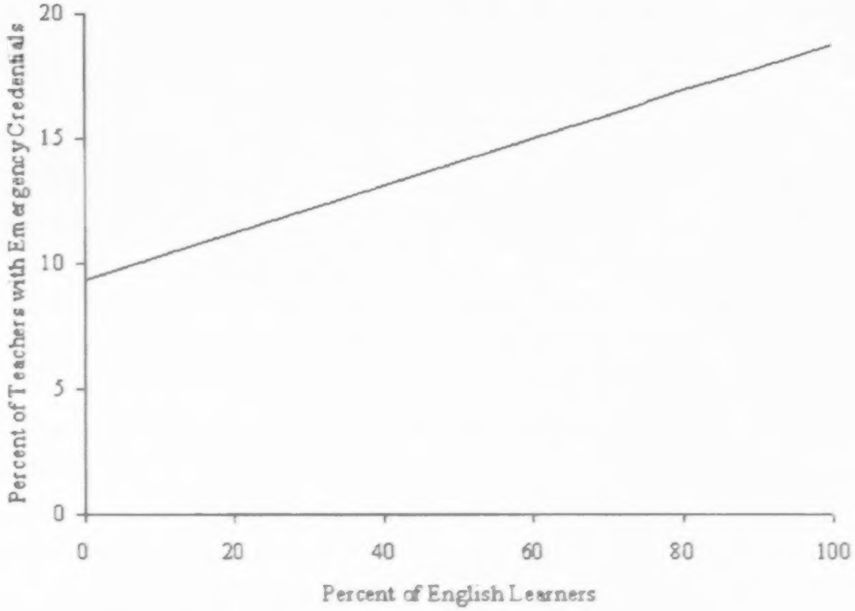
	Reading		Language	
	Actual Gains	Adjusted Gains	Actual Gains	Adjusted Gains

BCLAD	**1.8** (n=142)	**1.6** (n=142)	**4.1** (n=148)	**2.4** (n=148)
CLAD/LDS	**2.0** (n=32)	**2.7** (n=32)	**1.0** (n=34)	**0.4** (n=34)
SB1969	*	*	*	*
A Level@	**1.8** (n=155)	**1.6** (n=155)	**0.3** (n=155)	**-1.5** (n=155)
No Authorization	**-2.4** (n=74)	**-2.9** (n=74)	**0.5** (n-93)	**-1.8** (n=93)

*Actual and adjusted gains were not reported here due to the small sample size.
@ LAUSD certifies language competencies of its teachers if they do not already hold a BCLAD; A Level indicates fluent bilingual.
Source: Hayes & Salazar (2001), page 36

Whereas 14 percent of teachers statewide were not fully credentialed, 25 percent of teachers of ELs were not fully certified (Rumberger, 2002). Figure 5 shows that as the concentration of ELs in a California school increases, so too does the percentage of teachers holding emergency credentials. Inasmuch as Figure 5 holds poverty constant, we would expect to see a flat line if the discrepancy in credentialed teachers were purely a function of poverty. These data show that English learners are significantly less likely to have a fully credentialed teacher than other low- income non-EL students. We will demonstrate that this is largely a problem of uneven *distribution* of qualified teachers among California's schools and classrooms.

Figure 5. The Relationship between the Percent of English Learners and the Percent of Teachers with Emergency Credentials, Holding Constant the Percent of Students on Free or Reduced Lunch, California Schools, 1999-2000



Note: Relationship estimated from the regression equation: 3.553 + .119*LUNCH + .095*ELL (N=6039), with LUNCH = 48.6 (sample mean).
Source: 1999-2000 API Growth Data File. Retrieved October 4, 2000 from WWW: http://api.cde.ca.gov/datafiles.html

Authorizations to Teach English learners

The current state of the art of teaching EL students employs three central methodologies for English learner instruction. The first strategy, specially designed academic instruction in English (SDAIE), is defined as "a set of systematic instructional strategies designed to make grade-level and advanced academic curriculum comprehensible to English learners with intermediate English language proficiency" (California Commission on Teacher Credentialing, 2001, p. 26). Another means of teaching EL students is through their primary language. Although the principal goal is to provide access to the core curriculum, in reality, this involves a continuum of strategies, from using the student's primary language solely for clarification of concepts presented in English to actually providing academic instruction in the primary language. A third strategy is English language development (ELD). It is "systematic" instruction of English language that is designed to (1) promote the acquisition of English-listening, speaking and reading and writing skills—by students whose primary language is other than English, and (2) provide English language skills at a level that will enable equitable access to the core curriculum for English learners once they are presented with academic content. (CTC, 2001, p. 3).

The California Commission on Teacher Credentialing (CTC) issues basically two EL credentials meant to ensure that teachers have skills in some or all of the above instructional strategies (see Table 4): the Bilingual, culture, language and development credential (BCLAD) and the Culture, language and development credential (CLAD). Many California teachers of English learners hold earlier versions of these specialized credentials that are generally considered equivalent and authorize them to teach English learners (Note 9).

Skills, knowledge, and instructional settings approved for each authorization

The most rigorous of the credentials, the bilingual, culture, language, and development (BCLAD) certification, requires that teachers have expertise in the areas of: 1) language structure, 2) methodology for first and second language development, and 3) cross-cultural competency. BCLAD teachers must also demonstrate competency in three additional spheres: 4) methodology for primary language instruction, 5 & 6) knowledge of a particular culture and language of emphasis. Many BCLAD teachers earn their expertise through a Master's Degree program or through a credential program with an emphasis on teaching English learners infused throughout the program's coursework and field placements.

BCLAD authorization requires extra expertise because it authorizes teaching in all settings with English learners. These teachers are authorized in the various methods of EL instruction for conveying academic content and promoting English language proficiency including primary language methods, "specially designed academic instruction in English" (SDAIE), and English language development (ELD). As speakers of a second language, these teachers tend to be more sensitive to the issues surrounding the acquisition of a second language and can communicate with students and parents in at least one language other than English. Thus, these teachers have a variety of skills to address a range of EL students' educational needs.

The next most comprehensive authorization, the CLAD certificate or credential includes the first three skill areas required of the BCLAD teacher: 1) language structure, 2) methodology for first and second language development, and 3) cross-cultural competency. Expertise in these areas is gained through a set of four college courses—or by passing exams on this content. CLAD teachers should have some experience of learning a second language but are not required to have a command of that language or culture that is required for BCLAD certification. CLAD holders are authorized to teach subject matter to EL students using SDAIE and other English language methods, and to teach English language development.

Staffing EL classrooms with BCLAD or CLAD teachers allows English learners to remain in self-contained classrooms. Classrooms without CLAD or BCLAD teachers may require that EL students be removed for ELD (or academic support), so called pull-out instruction (Brisk, 1998). Despite being ubiquitous in English learner education, pull out instruction has been found to be among the least successful of instructional strategies for these students (Lucas, 1997; Ovando & Collier, 1998). Although BCLAD certification is the most comprehensive, it is also the rarest: Only 5% of California teachers who instruct English learners have a full credential with BCLAD authorization (UC LMRI, 2003).

"Teacher in training" Status

According to the California Commission on Teacher Credentialing, the most widely used option to teach English learners is the "teachers in training" status, which does not require any certification. Rather, teachers in training are permitted by the California Department of Education to teach English learners using ELD and SDAIE methods based upon a mere agreement to obtain the requisite training for certification within two or three years. Teachers in EL classrooms who sign agreements that they are participating in or will obtain the requisite training are conditionally allowed to continue in their positions by the CDE. Unlike the various other certifications offered, the teachers in training status is not monitored by the Commission on Teacher Credentialing. Rather, this status was developed by the California Department of Education as part of a "plan to remedy" the shortage of teachers certified to teach ELs in school districts that were found by the Coordinated Compliance Review (CCR) to be out of compliance with matters concerning English learners. However, instead of remedying the shortage of certified teachers, the CDE's re-labeling of untrained teachers has largely reinforced the status quo. Thus far, CDE monitoring and enforcement of these agreements has not resulted in any substantial reduction of the numbers of "teachers in training".

Supply of EL Authorized Teachers in California

To determine whether the problem is one of sufficient numbers of teachers qualified to teach English learners, or simply a maldistribution of qualified teachers, we analyzed figures from the California Basic Educational Data System (CBEDS) for the year 1999- 2000. CBEDS conducts an annual survey of every professional educator working in the public school system. Teachers are asked to indicate the type of California teaching credential they hold, including whether it is a "full" credential or an "emergency" credential. Teachers are also asked to indicate all the areas that their credential authorizes them to teach. We identified all teachers who indicated that they were authorized to teach in bilingual, English language development, or specially designed academic instruction in English (SDAIE) classes. We then compared the number of teachers with such special authorization to teach ELs with the number of EL students, both in the state as a whole and in each school that enrolled English learners. We also compared these figures with data on students who were not English learners and teachers without authorization to teach English learners.

Table 2. California Students and Teachers by Language Background, 1999-2000

	English Learner	Other	Total
Students	1,480,406	4,471,206	5,951,612
Teachers, *including* emergency permits/waivers	79,215a	212,840	292,055
Students per teacher	18.7	21.0	20.4
Teachers per 100 students	5.4	4.8	4.9
Teachers *excluding* emergency permits/waivers	75,687 a	175,781	251,468
Students per teacher	19.6	25.4	23.7
Teachers per 100 students	5.1	3.9	4.2
Fully credentialed bilingual/ELD teachers	69,305 b		
Student per teacher	21.4		
Teachers per 100 students	4.7		
Fully credentialed bilingual teachers	26,539 c		
Students per teacher	55.8		
Teachers per 100 students	1.8		

a. Teachers authorized in any way to teach bilingual education, English Language Development, or specially designed academic instruction in English (SDAIE), including those with SB1969 authorizations.
b. Teachers authorized to teach bilingual education or English Language Development.
c. Teachers authorized to teach bilingual education.

SOURCE: 1999 CBEDS and 2000 Language Census.

The figures in Table 2 show that in the state as a whole, there were almost 6 million students and almost 300,000 teachers, which represents 20 students per teacher or 5 teachers per 100 students. There were also about 1.5 million English learners and about 79,000 "EL" teachers, that is, those with some kind of special authorization (BCLAD, CLAD, 1969/395) to teach them through the primary language and/or ELD, and/or SDAIE. Ignoring for the moment whether all of these authorizations are adequate to the task of teaching English learners, this represents about 19 students per EL teacher or more than five EL teachers per 100 EL students. These figures suggest that there are slightly more teachers with some specialized preparation per EL student in the state than the statewide student/teacher ratio. The same conclusion can be drawn if a similar analysis is done with only teachers who are fully authorized to teach English learners: there are actually more fully authorized EL teachers in the state per EL student than there are fully credentialed (non-EL) teachers per non-EL student. However, the language census data indicating how many EL students are actually taught by authorized CLAD or BCLAD teachers paints a somewhat different picture. These data indicate a statewide average of only 4.2 CTC authorized EL teachers per 100 English language learners (California Department of Education, Education Demographics Office, Spring 1999 Language Census).

Comparing the numbers of teachers with the most rigorous training to teach English learners, those with BCLAD, bilingual specialist, or BCC credentials, the picture changes dramatically. Based on the same procedure as above, there are only 1.9 fully credentialed BCLAD equivalent teachers (*i.e.*, those with the most comprehensive credential) for every 100 EL students versus 3.8 fully credentialed teachers per 100 non-EL students, or half as many. Under this scenario, the state would need another 26,000 teachers with the most comprehensive credentials to reach the same proportion as for non-EL students taught by teachers with the most comprehensive training.

Distribution of EL Teachers in California

While this statewide picture suggests that there are sufficient numbers of EL teachers with at least some authorization to teach English learners, it does not indicate how those teachers are distributed among schools. To investigate this issue, we classified schools based on the number of fully credentialed EL teachers they had for every 100 EL students. We divided schools into four groups: (1) schools with no EL teachers, (2) schools with a ratio of fewer than 2.5 fully credentialed EL teachers per 100 EL students--half the state average, (3) schools with a ratio between 2.5 and 7.5, and (4) schools with a ratio of more than 7.5--50 percent above the state average. We then computed how many schools were in each category and how many EL students attended those schools (Table 3).

Table 3. Number of English Learners by EL Teacher/Student Categories and Level, 1999-2000

Fully Authorized EL teachers per 100 EL students	Elementary		Middle		High	
	Number	Percent	Number	Percent	Number	Percent
No EL teachers	18,689	1.9	5,703	2.4	1,675	0.7
Fewer than 2.5	193,205	19.7	81,954	35.3	74,119	31.3
2.5 to 7.5	610,629	62.3	120,153	51.7	132,402	55.8
Greater than 7.5	157,331	16.1	24,671	10.6	28,933	12.2
Total	979,854	100.0	232,481	100.0	237,129	100.0

Source: 1999 CBEDS and 2000 Language Census

At the elementary level, more than 200,000 English learners--20 percent of the total--attend schools with 2.5 or fewer EL teachers per 100 English language learners. At the middle school level, more than 85,000 ELs attend such schools-- almost 38 percent of the total. At the high school level, more than 75,000 attend schools with such low numbers of qualified EL teachers--almost one-third of all high-school EL students. Counting English learners who attend other types of schools (e.g., alternative, continuation, etc.), more than 390,000 English learners in California--one out of every four--attends a school with fewer than half the state average of teachers with specialized authorizations to teach them.

Another indication of the shortage of teachers with the appropriate training to teach English learners is revealed from an analysis of the 2000 Class Size Reduction (CSR) teacher survey (Stecher & Bohrnstedt, 2002). According to those data, 37 percent of all teachers who taught grades 1-4 in 2000 held a CLAD credential, 10 percent held a BCLAD credential, and 45 percent held either a CLAD or BCLAD (see Table 4). In general, the higher the concentration of English learners in the classroom, the higher the proportion of teachers who held at least some authorization to teach them. Yet among classrooms where a majority of students are English learners, only about half of the teachers held an appropriate EL credential. Using data on the proportion of English learners in each type of classroom, we estimate that only 53 percent of all English learners enrolled in grades 1-4 in California in the 1999-2000 school year were taught by a teacher with any specialized training to teach them (Note 11). If we assume that teachers with BCLAD credentials have the most appropriate training, only 22 percent of all English learners enrolled in grades 1-4 had such a teacher in 2000.

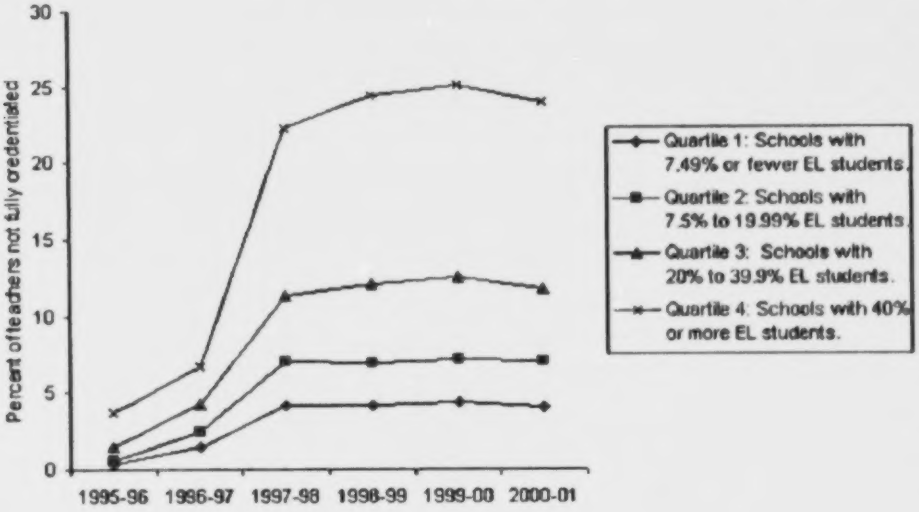
Table 4. Percent of Teachers in Grades 1-4 with CLAD and BCLAD Credentials By Classroom Concentration of English Learners, 2000

Percent English Learners in the classroom	Percent of all English Learners	CLAD	BCLAD	CLAD or BCLAD
0	0	25	2	27
1-25	17	47	3	47
26-50	20	46	11	54
51-100	63	29	30	54
Total	100	37	10	45

NOTE: Results are weighted.
Source: 2000 Class Size Reduction Teacher Survey (N=774).

Class size reduction had some largely unanticipated consequences for EL students because of the relative concentration of English learners in the state's poorest schools. The migration of credentialed teachers away from these schools to those in more affluent areas with better working conditions was a significant feature of the class size reduction initiative in California (Stecher & Bohrnstedt, 2002). For example, the percentage of teachers not fully credentialed in schools with the smallest proportion of English learners (less than 8 percent) only increased from .3 percent in 1995-96 to 4.0 percent in 2000-01 (see Figure 6). However, the percentage in schools with the greatest proportion of English learners (40 percent or more) increased from 3.7 percent to 23.9 percent over the same five-year period. As a result, schools with the most English learners benefited the least from class-size reduction, at least in terms of access to fully credentialed teachers.

Figure 6. Percentage of Public K-3 Teachers Not Fully Credentialed by School Quartiles of English Learners: 1995-96 to 2000-01



SOURCE: Brian M. Stecher and George W. Bohrnstedt (Eds.), *Class Size Reduction in California: Findings from 1999-00 and 2000-01* (Palo Alto: American Institutes for Research, 2002), Table B.17.

At the same time that EL students are less likely than others to have a qualified teacher, the challenges associated with teaching them are even greater than for the typical student. The large number of English learners who are immigrants frequently come from circumstances in which their early lives and education have been disrupted by war, loss or estrangement of family members, poverty, and residential mobility (Ruiz de Velasco & Fix, 2000; Olsen, 1998). As such, teachers must know how to intervene educationally with students whose personal and educational backgrounds are significantly different from the mainstream English-speaking student. Moreover, the age and grade placements of these students in U.S. schools often do not match their skill levels because of varying educational experiences in their countries of origin (Ruiz-de-Velasco & Fix, 2000).

Necessary Competencies of EL Teachers

Wong Fillmore and Snow's (2000) study, entitled *What Teachers Need to Know about Language*, outlines the critical knowledge base that teachers must have for language learning in order to effectively teach children who do not speak standard English. They note that teachers need to know the units of language and how they operate differently across languages and dialects. For example, knowing how tense and plurality are formed in the child's native language can help the teacher to uncover difficulties in English and facilitate learning for ELs. Wong Fillmore and Snow also argue that by knowing the fundamental characteristics of words in the primary language of the student, the teacher can facilitate more rapid acquisition of English vocabulary and word construction. They point out, for example, that if a teacher can explain that the suffix *idad* in Spanish has the same consistent meaning as *ity* in English, the student's vocabulary and word usage can be expanded significantly. These authors also assert that teachers must understand the norms for language usage in the primary culture of the student in order to know how to encourage English learners in their acquisition of English. Another critical competency that Wong Fillmore and Snow argue teachers must have is a clear understanding of what constitutes *academic English* and how to support the acquisition of this particular form of the language for English learners. Academic English is the language of texts and often of tests, and it is not normally acquired in the course of conversation outside of academic contexts. For students who are not likely to "absorb" this form of English discourse in their homes or communities, it must be explicitly taught.

Of course, one of the most controversial of all topics in education is the best method for teaching reading. Many experts argue that there is no single best method. Rather there are a number of strategies that are more or less effective with different students at different points in the process of learning (Snow, Burns & Griffin, 1998), and it requires the expertise of a well-trained teacher to know which strategy to use when, and with which children. There is even less agreement, however, on how best method to teach English learners to read in a language they do not understand–English. The National Research Council (Snow, Burns & Griffin, 1998) concluded that if reading instruction is not done initially in the primary language of the child, then educators should consider delaying it until English is acquired. This points out the degree to which the field continues to depend upon the skills of highly qualified teachers to make judgments about how best to teach reading to English learners. There simply are no "tried and true" strategies for teaching children to learn to read in a language they do not understand, and it is a vexing problem even to the experts in the field

Finally, there are significant issues associated with the cultural backgrounds of immigrant and non-English speaking students that bear on how they learn. Wong Fillmore and Snow (2000) argue that a critical role for all teachers is to socialize students to the demands of schooling. Certainly, this is made more difficult if the teacher does not understand both the cultural and linguistic norms of the students he or she is teaching. They point out that in correcting some students, or encouraging others to participate in linguistically-based activities in the classroom, teachers may inadvertently squelch the motivation of English learners to participate at all. Without understanding the cultural and linguistic expectations of these students' communities, teachers can undermine their students' learning by failing to acknowledge culturally appropriate behavior. For example, many teachers reward students for questioning behaviors and active participation in discussion, but some immigrant students are socialized to believe that such behavior is inappropriate in the classroom (Olsen, 1997).

Teachers themselves have cited their need for greater expertise in working with EL students. In a survey of all 1999-2000 graduates of teacher credential programs in the California State University system (total of 10,512) one fourth responded that they felt they were only "somewhat prepared" or "not at all prepared" to teach English learners (Office of the Chancellor, 2002). We note that these are the "cream of the crop" of teachers of English learners—those who have completed a full credential and in most cases have training at least at the level of the CLAD (Culture, Language and Development preparation (CLAD) credential (Note 12).

Many teachers of English learners also report difficulty in communicating with the parents of their students. In a recent Harris survey (Note 13), 23 percent of teachers of English learners reported that they had a hard time communicating with their English learners' parents about their children's educational progress and needs (Table 5). Not surprisingly, teachers with no special preparation to teach English learners were more likely to report difficulty, while most teachers with BCLAD credentials reported that they were able to communicate with their students' parents. It is also notable that in the Hayes et al. (2002) study of the LAUSD implementation of Proposition 227, the largest concern noted by non-English speaking parents was lack of communication with teachers.

Table 5. Percent of teachers with EL students in their classes who reported difficulty in communicating with parents by teacher credential, January 2002

Teacher credential	Percent reporting difficulty
CLAD (or equivalent)	25
BCLAD (or equivalent)	7
SB-1969/395 Certificate	10
None	30
Total	23

NOTE: Results exclude respondents who did not answer question or answered "not sure." Results are weighted. Overall differences are statistically significant at .05 level or better.
Source: Harris Survey of a Cross-Section of California School Teachers, January 2002 (N=820).

A common critique of teacher preparation programs, both in California and elsewhere, is that the extant knowledge of how to teach English learners is not often incorporated into teacher preparation efforts (Wong-Fillmore & Snow, 2000; Reagan, 1997; Tomas Rivera Center, 1994; Milk, 1990). In effect, we know considerably more about how to prepare teachers than we act on in schools of education. This is generally viewed as a problem in translating research into practice. The reasons for this have been debated at great length in the education literature (cf. Cochran- Smith & Fries, 2001), but the only firm conclusion that can be drawn is that there is a clear disjunction between research and practice in teacher education. It is nowhere more painfully evident than in the preparation of teachers for English learners where it is commonly argued that the field lacks research-based methods, when in fact this is not the case (cf. August & Hakuta, 1997; Gándara & Maxwell-Jolly, 2000).

Working Conditions for EL Teachers

Given the opportunity, teachers vote with their feet for where they want to work, and school conditions appear to influence this vote. Recent research suggests that working conditions influence teachers' decisions about where to teach more than do salaries (Hanushek, Kain, & Rivkin, 2001; Loeb & Page, 2000). Data for California demonstrate this clearly. Table 6 demonstrates that the differences between conditions in schools with high and low concentrations of EL students are dramatic, even with respect to characteristics that would not intuitively seem to be related to the concentration of English learners.

However, it is evident that when working and learning conditions are poor, they affect the attitudes of staff, and no doubt the ability of the school to attract competent and amiable people to work there.

Table 6. Characteristics of the Environment of California Elementary Schools by EL Concentration, Spring 1999

	25% or less	Over 25%	Total
Problems in neighborhood where school is located:			
Selling or using drugs or excessive drinking in public	16	50	29
Gangs	32	77	50
Crime	24	77	45
School climate:			
Parents not active in programs	4	40	18
Problem with overcrowding	25	50	34

Note: Results are weighted (S2SAQW0). All column differences are statistically significant at .05 level or better.
SOURCE: ECLS base year data for California public elementary schools (N=69).

Given the difficult working conditions and the added demands placed on teachers of English learners, it would be expected that both training and guidance on how to address these challenges would be provided. The data, however, show otherwise. Teachers of English learners are largely left to fend for themselves with inadequate guidance, resources, and training.

(2) Inadequate professional development opportunities to help teachers address the instructional needs of English learners.

Although there are no studies that are able to demonstrate a causal relationship between student achievement outcomes and teacher professional development, the relationship between teacher preparation and student achievement (see, for example, Darling- Hamond, 2002) suggests that such a relationship probably does exist. Moreover, a number of studies have demonstrated that good professional development increases teachers' sense of competence and provides them with tangible strategies for better meeting the needs of their students (Herman & Aguirre-Munoz, 2003; Herman, Goldschmidt, & Swigert, 2003). Given these findings, teacher professional development has been a cornerstone of many states' education reform plans, including California's. Yet, surprisingly little emphasis has been placed on the specialized needs of teachers of English learners.

The instructional demands placed on teachers of English learners are intense. They must provide instruction in English language development while simultaneously or sequentially attempting to ensure access to the core curriculum. Yet data collected for California's Class Size Reduction Study (Stecher & Bohrnstedt, 2000) show that even where teachers are teaching a majority of English learners, the professional development they receive that is dedicated to helping them instruct these students is minimal. The percent of professional development time that teachers reported focusing on the instruction of English learners in 1999-2000 ranged from 3 to 10 percent with a mean of only 7 percent (Table 7).

Table 7. Professional Development of Teachers in Grades 1-4 by Classroom Concentration of English Learners, 1999-2000

Percent English Learners in the classroom	Mean number of days	Mean number of hours[a]	Percent of hours on teaching English Learners
0	3.5	28	3
1-25	3.6	30	8
26-50	3.3	32	9
51-100	3.8	35	10
Total	3.6	31	7

Number of hours estimated by recoding responses (8 hours or less = 4 hours; more than 8 = 12 hours).
Note: Results are weighted.
Source: 2000 Class Size Reduction Teacher Survey (N=774).

These data are corroborated by several other recent studies. Hayes and Salazar (2001), in their study of 177 classrooms in the LAUSD, noted that teachers discussed "the problematic lack of resources and training to assist them to provide quality services to ELLs (p23)." A report on the results of a California

Department of Education (CDE, 1999) survey of every California school district during the first year of Proposition 227 implementation showed that professional development to help teachers with English learner instruction was one of the most significant unmet needs in the aftermath of the passage of the proposition. The later, more ambitious, study of the implementation of Proposition 227 being conducted by American Institutes for Research (Parrish, et al., 2001, 2002) likewise reports a similar theme emerging from their investigation. The study documents a significant lack of guidance from the state about the nature of the instruction that should occur in the Structured English Immersion classrooms, and as a result, "teachers were not provided appropriate materials or guidance on how to use materials appropriately" (Parrish, et al., 2001, p. 36). Again, in the most recent report of this five-year study, researchers concluded that, "Barriers to the implementation of the Proposition include insufficient guidance for implementing regulations in the law; confusion over what the law requires and allows; and lack of clear operational definition for the various instructional approaches for EL students. In particular, educators lack clarity on what constitutes best practice within structured English immersion instruction" (Parrish, et al., 2002, p. ix).

The University of California has provided professional development for the state's teachers through its Professional Development Institutes (CPDIs). This is not the only professional development activity in the state, in fact, many districts sponsor extensive professional development programs, but it is the largest state-wide effort, with more than 45,000 teachers participating in these workshops in 2000-01. In that same year, a total of $50,866,000 was provided for this purpose. Of this amount, only $8,358,104 was earmarked for professional development in the area of English Language Development (Office of the President, University of California, 2002). This constituted about 16% of the professional development budget, although English learners constitute fully 25% of the students in the state, and as we have argued, are the most educationally deprived by their schools of all students. The AIR study found that only 18% of the teachers in their sample had even heard of the ELD CPDIs, and only 8% had attended one or more (Parrish, et al, 2002, p. IV-40), suggesting that relatively little is done to disseminate information about resources that may be available to teachers of English learners.

3. Inequitable access to appropriate assessment to measure EL achievement, gauge their learning needs, and hold the system accountable for their progress (Note14)

While English learners must be incorporated into a state accountability system in order to insure that their educational needs are being met, the current system is of little value for monitoring their academic progress.

English Language Testing of the Academic Progress of English Learners

According to the National Academy of Sciences, "when students are not proficient in the language of the assessment (English), their scores on a test in English will not accurately reflect their knowledge of the subject being assessed" (NRC, 1999, p. 214). Therefore such assessments provide neither accurate data for accountability purposes, nor do they help teachers to enhance their instruction. As the National Research Council noted, " if a student is not proficient in the language of the test, her performance is likely to be affected by construct-irrelevant variance –that is, her test score is likely to underestimate her knowledge of the subject being tested" (NRC, 1999, p. 225). These tests can, moreover, have serious negative effects on the schooling of English learners in at least two ways: (1) positive changes in test scores over time can give the inaccurate impression that students have gained subject matter knowledge when, in fact, they may have simply gained proficiency in English. This misperception that EL students are making academic progress can lead schools to continue providing a curriculum that fails to emphasize comprehensible subject matter. (2) On the other hand, consistently low scores on tests can lead educators to believe that students need low-level or remedial education, when in fact, they may have mastered the curriculum in another language, but are unable to express these competencies through an English language test.

The current state accountability practice for English learners is as follows:

- All EL students in Grades 2-11 must take the Stanford 9 (SAT 9) (Note 15), a nationally norm-referenced test in reading and math (and science and social studies in the higher grades) administered in English, unless parents or a guardian provides a written request for a waiver.
- Only "ELLs who have been in the district for 12 months or more may not use nonstandard accommodations unless they have individualized education plans (IEPs) or Section 504 plans that include accommodations." (Note 16)
- English learners who have been in a district for less than one year (except for entering ninth graders in high school districts as of 2000) are excluded from the Academic Performance Index (API). (Note 17)
- The API is used to measure each school's performance based on student test scores. Through the Governor's Performance Award (GPA) Program, there are monetary and non- monetary awards based on positive changes in the API. In addition, through the Immediate Intervention/Underperforming Schools Program (II/USP), failing schools are identified for local and state intervention to improve programs
- "Spanish-speaking English language learners who had been enrolled in California public schools

less than 12 months when testing began [are] required to take the SABE/2 in addition to taking the Stanford 9..." (Note 18)

- Finally, the CAHSEE is the gatekeeper of graduation that all students, including English learners and other students with exceptional needs, must pass in order to receive a high school diploma (as of 2007).

The exclusive reliance on an English-language norm-referenced achievement test for EL students is inappropriate for these students (Note 19) and violates several standards established by the authoritative AERA/APA/NCME Standards for Educational and Psychological Testing. (Note 20)

Research on second language acquisition shows that it takes English learners on average between four to seven years to meet various standards of English proficiency (Hakuta, Butler & Witt, 2000).The burden is on the state to demonstrate that test scores for English learners who have been in the United States for less than four years are valid, yet the state has not made any attempt to obtain information to shed light on this question (Note 21).

The only cautionary statement by the CDE on the interpretation of standardized test scores appears on a web page and says: "Since the Stanford 9 norming sample was representative of the United States as a whole, it does not necessarily match California's student population." (Note22) There is no acknowledgement such as that of the San Diego Education Association that explicitly states that the California population is vastly different from the norming group: "The norming sample, while representative of the nation, does not reflect the huge diversity of California's student population. For example, 40.5% of California's students are Hispanic, but only 9.6% are in the Stanford 9 norming sample. While 24.6% of the state's students are of limited English proficiency, only 1.8% are in the sample." (Note 23) Since the test scores are reported with respect to the national percentile rank (NPR), failure to issue an explicit warning with respect to Hispanics and to English learners is a clear violation of this standard.

The state requires the collection of data using a comparable test in Spanish, the SABE/2. However, the state gives no guidance on how this information might be used to augment information from SAT 9. Indeed, while requiring SABE/2 for Spanish speaking students who have been in the U.S. for less than 12 months, the Department of Education explicitly rejects the use of SABE/2 in its Academic Performance Index (API) and does not monitor its administration. (Note 24) Among the reasons for this according to the Superintendent's Advisory Committee for the Public Schools Accountability Act of 1999 is that "SABE/2 is not aligned with state content standards", leading to "consensus in the API subcommittee to not include the SABE/2 in the 1999-2000 API." (Note 25) That SABE/2 is not aligned with the state content standards is uncontestable. But neither was the SAT 9 at the time the API system was developed. More recently, the SAT 9 has been augmented with new and revised items to bring it into alignment with state standards, and in fact, a new test is being developed that will be in even greater alignment. The same could be done with the SABE/2 or another similar achievement test in Spanish. Given the importance of API as a policy instrument in driving rewards and sanctions for school academic performance, it is indefensible that the state makes no provisions for the use of information from a native language test, indeed one that the state requires (on paper only) students to take. This point is made particularly salient by the discrepancies between English learner scores on the SABE/2 and on the SAT 9. Well over half (59%) of all 4th graders taking the SABE/2 reading test in 2001 scored at or above the 50th percentile on this test, which was normed on a Spanish-speaking population. In contrast, only 15 percent of English learners in the 4th grade performed as well on the SAT 9 (Note 26). The tests are not strictly comparable, but the discrepancies raise serious questions about the appropriateness of current practice of educational planning based on clearly faulty and incomplete information about what EL students know and need to know.

California English Language Development Test (CELDT)

The CELDT test of English language development was developed to make it possible for educators to assess the level of English proficiency of their English learners. Administered for the first time in 2001, the test, developed by CTB-McGraw Hill, and designed to assess English learners in grades 1 through 12 on reading, writing, comprehension, and oral English along a continuum of five levels, from Beginning (1) to Advanced (5). The purpose of the test is to monitor the growth of students' English skills over time and to provide a single statewide measure useable for both program placement and reclassification to the status of English proficient. Prior to the development of the CELDT, school districts used a variety of different tests which did not align with each other and provided no consistent basis for monitoring student progress (Gándara & Merino, 1994). While the CELDT was warmly received by educators of English learners as a potentially useful assessment tool, it suffers from at least two major limitations. It was so time-consuming to administer that it placed a significant burden on schools. In addition, scores were not available in a timely enough fashion to allow educators to use the information for program purposes (Note 27). As a result of these limitations, the CELDT has been revised by CTB-McGraw Hill. The restructured CELDT, form C, is being administered at the time of this writing. To what extent this makes it impossible to compare scores from 2001 to 2003 is not yet known.

California High School Exit Exam

EL students are much less likely to pass the High School Exit Exam (CAHSEE) than are English speakers (Note 28). Students with exceptional needs as defined in Title 20 of federal law may take the exams with accommodations to meet their special needs. However, English learners do not have exceptional needs according to this definition and do not therefore qualify for accommodations in the state of California. The law does allow for districts to defer the requirement that students pass the exam until the pupil has completed six months of instruction in reading, writing, and comprehension in the English language. Nonetheless, no student, including those who are still classified as English learners, will receive a high school diploma without passing the exit examination in English.

An important feature of the law authorizing the CAHSEE is a requirement that the exam have curricular and instructional validity:

> (2) "Curricular validity" means that the examination tests for content found in the instructional textbooks. For the purposes of this section, any textbook or other instructional material adopted pursuant to this code and consistent with the state's adopted curriculum frameworks shall be deemed to satisfy this definition.
>
> (3) "Instructional validity" means that the examination is consistent with what is expected to be taught. For the purposes of this section, instruction that is consistent with the state's adopted curriculum frameworks for the subjects tested shall be deemed to satisfy this definition (Education Code Section 60850, f, 2 & 3).

The evidence that EL high school students do not receive the same instruction or have access to the same range of courses as their English-speaking peers puts in serious question the curricular validity of these tests for English learners. Moreover, evaluation of the early administrations of the test found that passing rates on the math exam are significantly correlated with completion of Algebra I (Wise, et al., 2002). Yet, EL students are often on a high school trajectory of ELD and basic classes that does not include algebra (Callahan, 2003). Furthermore, the condition of these students as English learners—students who by definition do not have the same level of understanding of all- English instruction as fluent English proficient students—raises questions about the instructional validity of the exam. This is particularly true for EL students in classrooms with teachers who do not have special certification or preparation in English learner teaching strategies. Unfortunately, the dearth of such prepared teachers (discussed in Section 2) is even greater at the secondary than the elementary school level.

Accommodations

When English tests are used to assess English learners, it is common practice in many states to use accommodations. Examples of test accommodations include using a parallel form of the same test content in the native language, administering the test in small groups, repeating directions, having a person familiar with the child's language and culture give the test, giving more time breaks, reading questions aloud in English, translating directions, extending the session over multiple days, simplifying directions, and using word lists or dictionaries (National Research Council, 1999: 218). California, however, does not allow accommodations for those EL students who have been here for over one year. CDE guidelines state:

> English Language Learners may use nonstandard accommodations only if the local board of education adopts a policy before testing begins that includes the criteria each school is to use to identify ELLs eligible to use nonstandard accommodations. ... After the policy is adopted English Language Learners who will have been enrolled in the district less than 12 months when testing begins may use any of the nonstandard accommodations including having the directions translated and using bilingual dictionaries. ELLs who have been in the district for 12 months or more may not use nonstandard accommodations unless they have IEPs or Section 504 Plans that include accommodations." (Note 29) Yet since the API index counts only those English learners who have been here for over one year, the API does not include assessment results for students who were allowed by a local school board to use accommodation practices.

The need for making accommodations available by at least allowing EL students additional time is clear from inspection of the data on the number of items and the time allotted, according to a table available on the CDE website (Note 30). For example, the reading vocabulary section of the test, at each grade level, has 30 items given in an allotment of 20 minutes and for reading comprehension there are 54 items to be completed in 50 minutes for most grades. This pattern is also found in the mathematics items. That is to say, the speed of the test is less than one minute per item. While this may be sufficient for native speakers of English, this is hardly sufficient for most English Learners, particularly given that tests such as the SAT 9 purport to be tests of achievement (or "power"), not of speed.

(4) Inadequate instructional time to accomplish learning goals

There is a significant body of research that shows a clear relationship between increased time engaged in academic tasks and increased achievement. Carroll (1963) devised the classic model that showed learning is a function of the amount of time needed to learn something divided by the amount of time allotted to learn it. Others have attempted to make sense of the different ways that time can be used productively in learning. Berliner (1990) argues that "time on task" is different from "academic learning time", with the latter resulting in greater achievement gains than the former. Karweit (1989) has shown that "engaged time" on task, which is akin to academic learning time, is more important than simply the time allotted, as in the Carroll model. All of these models, however, suggest that there is a relationship between time and learning, and that learning increases when students are optimally engaged in learning activities for greater amounts of time. Notwithstanding the importance of time for learning, there are many ways in which English learners experience less time on academic tasks than other students:

- With the passage of Proposition 227, English learners who enroll in a California school for the first time must remain in a structured English immersion program for at least 30 days before being assigned to a permanent classroom. In a recent study of schools implementing the proposition, many teachers complained that they did not know what to do with students during this interim period and that a great deal of instructional time was lost trying to accommodate students who would not be continuing on in the same classroom. Particularly where parents had sought a waiver to have their child attend a bilingual classroom, teachers reported not knowing how to instruct these students. They lacked the necessary curricula and materials for the 30 days of all-English instruction before they began what would be their bilingual program for the remainder of the school year (Gándara et al, 2000).
- A common way that elementary schools organize instruction for English learners is to take them out of their regular classes for English language development. This strategy has been demonstrated to create further inequities in the education of "pulled out" students because they miss the regular classroom instruction (Cornell, 1995; Fleishman & Hopstock, 1993; Anstrom, 1997). Nevertheless, the practice continues to be relatively routine for English learners. There is generally no opportunity for students to acquire the instruction they have missed during the pull out period (Lucas, 1997; Ovando & Collier, 1998).
- In secondary schools English learners are often assigned to multiple periods of English as a Second Language (ESL) classes while other students are taking a full complement of academic courses. Commonly, when not enough courses are available in either SDAIE or other formats, students are given shortened day schedules, resulting in significantly less time devoted to academic instruction (Olsen, 1997).
- Classrooms with large numbers of English learners also have fewer assistants in them to help the teacher provide individualized time for the students. Table 8 shows the number and types of person hours devoted to classrooms by percent EL.

Table 8. Hours of Assistance on Instructional Activities in Classrooms of Teachers in Grades 1-4 by Type and Classroom Concentration of English Learners, 1999-2000 (Mean hours)

Percent English Learners in the classroom	Regular aides	Special education aides	LEP or bilingual aides	Parents or adults	Students	Other specialists	Total
0	3	2	<1	4	1	1	11
1-25	3	1	<1	2	1	1	8
26-50	2	<1	2	1	1	<1	7
51-100	3	<1	2	<1	1	1	7
Total	2	1	1	2	1	1	8

NOTE: Results are weighted.
SOURCE: 2000 Class Size Reduction Teacher Survey (N=774).

While the district is apt to provide more bilingual aide time for classrooms with high percentages of English learners, there is significantly less time spent in these classrooms by parents or other adults. The result is that classrooms with no or few English learners enjoy more adult time in the classroom, which means that more of these children will receive individualized instructional attention, exacerbating the gaps in instruction and achievement outcomes between English learners and English speakers. While it is not necessarily the school's or the district's "fault" that some schools enjoy more parent participation, it is a fact that must be considered in distributing resources among schools. Furthermore, when EL students are taught by bilingual teachers these teachers are provided with much less paraprofessional assistance than their non-bilingual colleagues. In the view of many teachers this constitutes "penalizing" the most prepared teachers, and their students, for their extra expertise.

English learners in California are also more likely to be assigned to multi-track year round schools designed to accommodate more students on a campus. The year round plan that accommodates the most students is Concept 6, a schedule in which students attend school for only 163 days per year, instead of the 180 mandated by state law (Note 31). As Table 9 shows, English learners comprise fully half of the students assigned to Concept 6 schools. Students on the Concept 6 calendar attend school for 4 months twice a year, with two month breaks in between. This provides English learners less time to assimilate critical academic material and to be exposed to English language models. Just as important, however, is the loss of learning that occurs with two months breaks in school every 4 months. A significant body of research has now established that low income children (and English learners) are more disadvantaged by these lengthy breaks from school than middle income children. There is a demonstrably negative effect on their achievement (Cooper et al., 1996). Thus, the very students who need the most exposure to schooling, to English language models, and to opportunities to "catch up" to their English speaking peers are more likely to be assigned to school calendars that provide them with fewer school days than other students and less exposure to English in a school setting.

Table 9. Distribution Characteristics of California Schools, 2001 Percent English Language Learner Enrollment by School Calendar

		Calendar		
Measure	Statistic	Traditional/ Single-Track	Multi-Track Not Concept 6	Multi-Track Concept 6
Percent English Learners	Mean	21	36	51
	Median	15	35	53
	Number of Schools	5,913	735	221

Source: California Department of Education, Policy and Evaluation Division (http://cdedata.com.hosting.pacbell.net/api2001base/dbapi01b.zip) and School Facilities Planning Division (http://www.cde.ca.gov/facilities/yearround/direct00.htm)

(5) Inequitable access to instructional materials and curriculum

All students need appropriate instructional materials. While some might argue that textbooks and other appropriate learning materials are not essential to learning, Oakes and Saunders (2002) have argued cogently that the preponderance of research evidence demonstrates a clear link between appropriate materials and curriculum and student academic outcome. English learners, however, need additional instructional materials in two areas. First, all English learners need developmentally appropriate materials to learn English and to master English Language Development standards. Second, English learners receiving primary language instruction need appropriate materials in their native language. However, the evidence suggests that many are not gaining access to such materials. In the second year report of the AIR study, researchers report that 75% of the teachers surveyed said they "use the same textbooks for my English learner and English only students" and fewer than half (46%) reported using any supplementary materials for EL students (Parrish, et al., 2002, p. IV-34). This raises the question of how much EL students can be expected to learn without materials adapted to their linguistic needs. It is not particularly surprising then that only 40.9% of teachers report they are "able to cover as much material with EL students as with EO students" (Parrish, et al., 2002, p. IV-35). There is ample evidence in the research literature that when students cover less material than their peers, their skills decline relative to other students and they are prone to be placed in low academic groupings or tracks where educational opportunities are limited (Barr & Dreeben, 1983; Oakes, 1985; Goodlad, 1984; Gamoran, 1992).

The quality of instructional materials appears to differ by concentration of English learners in the school as well. Data from the Harris survey show that teachers with high percentages of English learners are less likely than teachers with low percentages of English learners to have access to textbooks and instructional materials, in general, and materials needed by English learners in particular. Almost half of teachers with high percentages of English learners report that the textbooks and instructional materials at their schools were only fair or poor compared to 29 percent of teachers with low percentages of English learners (Table 10). Teachers with high percentages of English learners were also almost twice as likely as teachers with low percentages of English learners to report that the availability of computers and other technology was only fair or poor. Moreover, almost two-thirds of teachers with high percentages of ELs in their classes reported not enough or no reading materials in the home language of their children and more than one quarter reported that they did not have any or enough reading materials at students reading levels in English.

Table 10. Condition and Availability of Instructional Materials in California Schools by Percentage of English Learners in Teachers' Schools or Classrooms, January 2002 (percent of teachers reporting condition)

	School EL		
	25% or less	Over 25%	Total
Reported by all teachers (N=1071)			
Textbooks and instructional materials are ONLY FAIR OR POOR	14	22	17
Availability of computers and other technology is ONLY FAIR OR POOR	26	40	31

	Classroom EL		
	30% or less	Over 30%	
Reported by teachers who have EL students in their classes (N=829)			
Not enough or no reading materials in home language of children	44	68	51
Not enough or no reading materials at students reading levels in English	19	29	22

NOTE: Results exclude respondents who did not answer question or answered "not sure." Results are weighted. All column differences are statistically significant at .05 level or better.
Source: Harris Survey of a Cross-Section of California School Teachers, January 2002.

Weak Curriculum

There is a significant body of research on tracking that connects the rigor of instruction to students' academic outcomes (Mickelson, 1999; Oakes, 1985; Page; 1990). The negative effects of low-track instruction are doubly detrimental for ELs who enter the classroom already at a linguistic disadvantage (Callahan, 2003; Harklau, 1994; Katz, 1999). There is a common perception that English learners are clustered in the early years of school, and so most attention is applied to students in this age group. However, about one-third of English learners in California are found in grades 7 – 12. And, these students are often shortchanged by their schools because of lack of appropriate coursework offerings or materials to support courses for English learners. In secondary schools, English learners are often assigned to multiple periods of ESL or ELD classes while other students are taking a full complement of academic courses. Commonly, when not enough courses are available in either SDAIE or other formats, students are given shortened day schedules, resulting in significantly less time devoted to academic instruction (Olsen, 1997).

We selected a random sample of transcripts of secondary English learners from two different northern California districts. In district #1, we compared a random sample of English learners with a random sample of English speaking students. For English only students (20) with GPAs from 1.6 to 4.1, 58% of their courses were college preparatory. For the English learners (8), with GPAs from 1.3 to 3.3 (this was the upper bound of the GPA range for English learners), 21% of their courses were college preparatory. The following are samples of English learner programs for the sophomore and senior years:

District #1

Saul (2 years in U.S., attended 9th grade in Mexico where he was in a college preparatory curriculum and took advanced mathematics courses) Sophomore year (2001):

Period 1: No class
Period 2: Language Development 1
Period 3: Language Development 2
Period 4: Native Spanish1
Period 5: U.S. History (in Spanish)
Period 6: Math A (general, low level)
Period 7: Weightlifting

(two courses meet college preparatory requirements: Spanish and U.S. History. No science is provided.)

Jose Luis (1 year in the U.S. Uneven academic history prior to immigration) Sophomore (2001):

Period 1: No class
Period 2: Language Development 1
Period 3: Language Development 1
Period 4: General Math (in English)
Period 5: Native Spanish 1
Period 6: Drawing 1
Period 7: No class

(One class prepares student for college requirements: Spanish. No science or social science offered. Student failed English only math because he could not understand the teacher.)

District #2

Marcos (Long term EL student, enrolled in California schools prior to entering high school).

Sophomore (2000):

Period 1: English 10 SDAIE
Period 2: World History SDAIE
Period 3: Pre Algebra A SDAIE
Period 4: Court Sports
Period 5: Integrated Science 2 SDAIE
Period 6: ELD 5

(Only two courses could be used to meet college preparatory requirements: World History and Integrated Science as an elective, not as a science course. Student never took a college preparatory science, math or English course through the junior year of high school.)

Marisela(Long term EL student, enrolled in California schools prior to entering high school)
Senior year (2002):

Period 1: Power English
Period 2: Weight training
Period 3: ELD 5C
Period 4: Business Math
Period 5: Consumer Foods
Period 6: Floral Design

(None of the student's courses meets college preparatory criteria. The student took no laboratory science or math beyond algebra. 1 which she failed and received no credit.)

These are students who have been attending California schools with caring administrators and school personnel, but the schools did not have the resources –human or otherwise—to provide an appropriate program of study for these students. They were selected randomly from among a pool of students like them for illustrative purposes, but they represent typical scenarios in many of California's high schools.

Because the state does not effectively monitor the quality of instruction that English learners receive, or the amount of time they spend in Structured English Immersion settings, we do not know to what extent the educational services provided for these students meet high standards of quality. We can guess at this figure, given the large numbers of unprepared teachers who teach them. It is worth noting, however, that more than 82,000 English learners in California receive *no special instruction whatsoever*. For some of these students this is based on parental request, but even this requires greater scrutiny. The AIR study of the implementation of Proposition 227 (Parrish, et al, 2002) noted that there remained a great deal of confusion among parents about what options existed for them, and that "in some cases, teachers are discouraged from discussing educational alternatives for students" (p. IV-41). In this environment, some parents are certainly making uninformed decisions about their children's educational program. The state has not monitored the extent to which schools and districts provide full disclosure to parents about the programs they may and do offer.

Over-placement in Special Education resulting in weak curriculum

The persistent and pervasive inequities in access to well- prepared teachers, school resources and facilities, appropriate assessment and time to accomplish learning goals result in large and growing gaps in achievement for English learners vis-à-vis their English speaking peers, and ultimately to misplacement into some special education classes. In the consent decree resulting from the ***Diana v California State Board of Education*** (U. S. D. C., ND, Cal.1970), a class action suit on behalf of English learners inappropriately placed in special education, the state agreed to the following:

- To test Mexican American children in their own language and in English
- To test them on the non-verbal sections of intelligence tests
- To re-test all Mexican American who are in Educable Mentally Retarded (EMR) classes using non-verbal sections of intelligence tests
- Develop and norm a test of IQ that reflects Mexican American culture
- Require school districts throughout the state that show a significant disparity between their overall district racial-ethnic representation and the racial-ethnic representation in their EMR classes to submit an explanation for the overrepresentation.

Thirty years hence the State of California has still not acted to implement the consent decree with respect to the development of appropriate assessment for English learners that could stem the over diagnosis and placement of these students in special education. Nor does California keep reliable data on the numbers of EL students in special education. About to be published is a study based on data from eleven school districts and over 700,000 students in the Los Angeles area for the 1998-99 school year. The researchers, Artiles and Rueda (in press) report that "ELs are over-represented in special education, particularly in specific learning disabilities (SLD) and language and speech impairment classes (SLI), especially at the secondary grade level where language support is minimal" (pg.2). Even more distressing is that, "highly vulnerable ELs (those who have low proficiency in both English and their primary language) are 1.5 times more likely to be diagnosed as Speech Impaired and Learning Disabled than their English speaking peers during the elementary school years. During the high school years, "highly vulnerable ELs" are twice as likely to be diagnosed as Mentally Retarded, Speech Impaired, and Learning Disabled. The state of being highly vulnerable –or having low proficiency in two languages—is often a product of inadequate instruction, just as proficiency in at least one language is the usual outcome of schooling and this is true for all children, regardless of their ability level. We know, for example, that many mentally disabled children acquire a reasonable proficiency in their primary language (Rueda, R. & Smith,1983; Whitaker, Rueda, & Prieto,1985). Table 11 shows that English learners and highly vulnerable English learners are significantly over-represented in special education programs in the sampled districts.

Table 11. Percent Students in Special Education, Elementary (K-5) & Secondary (6-12) Compared to Percent of Total School Population by language status and White (non EL), 11 Los Angeles Area School Districts

	White	Typical EL	Highly Vulnerable EL
Elementary	(9%)*	(28%)*	(22%)*
SLI	14%	17%	**48%****
SLD	13%	29%	**48%****
Secondary	(12%)*	(12%)*	(13%)*
MMR	14%	9%	**26%****
SLI	11%	10%	**27%****
SLD	9%	**17%****	**23%****

*p<.05; **p<.01
Source: Artiles & Rueda, in press

As was the case with the 1982 report by the National Academy of Sciences (Heller, Holtzman & Messick, 1982), an important finding is underscored by Artiles and Rueda: where there are few if any primary language support services offered, special education misdiagnosis and misplacement tends to occur. This is almost certainly related as well to the inequitable distribution of psychologists in the schools who can meet the assessment needs of English learners. The National Association of School Psychologists reports that only 160 out of all school psychologists in California report having bilingual competency. There are currently 1,949 school psychologists employed in California schools. If all of the bilingual psychologists were employed in the schools (which they almost certainly are not) then only 8% of psychologists would be bilingual and capable of conducting an assessment in a student's primary language. And, if all of these psychologists were assigned only to English learners, then 8% of the psychologists would be assessing 25% of the students.

Placement in special education, especially when it is not warranted, can have devastating effects on students' access to opportunities later in life. Evidence has existed for years documenting the massive rates of high school non-completion, underemployment, poverty, and adult marginalization of special education students after they leave high school (Guy, Hasazi, & Johnson, 1999). Placed in a special education track, it is unlikely for students to rejoin the mainstream. Robert Peckham, the presiding judge for the *Diana* case, summarized the evidence on the effectiveness of California's special education program, calling it a "dead-end educational program" (**Crawford v. Honig**, 1988).

(6) Inequitable access to adequate facilities

While it has been notably difficult to establish a firm link between the quality and condition of school facilities and the educational outcomes for students-- largely because the quality of school facilities is so highly correlated with wealth of the students and communities that schools serve – there is considerable consensus that it is difficult to both teach and learn in grossly inadequate facilities (Ortiz, 2002). Equally important, as we have already established, the conditions of schools are also related to teacher turnover. Teachers do not want to teach in dirty, dangerous, and uncomfortable conditions, and so they leave when they can. And, given that it is exceptionally difficult, if not impossible, to effect school reform without a stable base of teachers, school facilities certainly play an important, albeit indirect, role in student achievement.

Teachers of English learners are more apt than teachers of English speakers to respond that they do not have facilities that are conducive to teaching and learning. In the Harris survey close to half of teachers in schools with higher percentages of English learners reported the physical facilities at their schools were only fair or poor, compared to 26 percent of teachers in schools with low percentages of English learners (Table 12). Teachers in schools with high percentages of English learners were 50 percent more likely to report bathrooms that were not clean and open throughout the day and having seen evidence of cockroaches, rats, or mice.

Table 12. Condition of Facilities of California Schools by Percentage of English Learners in Teachers' Schools, January 2002 (percent of teachers reporting condition)

	25% or less	Over 25%	Total
The adequacy of the physical facilities is ONLY FAIR OR POOR	26	43	32
Bathrooms ARE NOT clean and open for throughout day.	13	23	17
HAVE seen evidence of cockroaches, rats, or mice in past year.	24	34	28

Note: Results exclude respondents who did not answer question or answered "not sure." Results are weighted. All column differences are statistically significant at .05 level or better.
Source: Harris Survey of a Cross-Section of California School Teachers, January 2002 (N=1071).

ECLS data show the same picture with regard to facilities. More than a third of principals in schools with higher concentrations of English learners reported that their classrooms were never or often not adequate, compared to 8 percent of principals with low concentration of EL students (Table 13). (Note 32)

Table 13. Characteristics of California Elementary School Facilities by EL Concentration, Spring1999

	25% or less	More than 25%	Total
Principal questionnaire responses:			
Classrooms never or often not adequate (Note 33)	8	35	19

Note: Results are weighted (S2SAQW0).
Source: ECLS base year data for California public elementary schools (N=69).

Such conditions not only make it more difficult to teacher English learners, they also make it difficult to retain teachers since, as we showed earlier, a considerable body of research finds that teachers are more likely to leave schools with poor working conditions.

(7) Intense segregation into schools and classrooms that place them at particularly high risk for educational failure

Research on desegregation has established that minority students who are schooled in desegregated settings tend to have better occupational outcomes and overall life chances (Wells & Crain, 1994; Crain & Strauss, 1985). Sociologists often explain this phenomenon as the impact of social capital –access to important social networks—on student outcomes (see for example, Stanton-Salazar, 1997). Therefore, one reason to be concerned about racial, ethnic, or linguistic isolation is the effect it has on limiting access to important social networks. However, a more immediate impact of linguistic isolation is the lack of appropriate English language models, which can result in both reduced opportunities to hear and interact with the language, and fewer opportunities to understand the ways in which the language is actually used in social and academic contexts (August & Hakuta, 1997). Both are important features in the development of high levels of linguistic skill.

We argue that the concentration of English learners in classrooms and schools in California compromises their opportunity to receive an education that is comparable in quality and scope to that of their non-EL peers because: (1) the lack of peer English language models limits the development of English; (2) the lack of models of children who are achieving at high or even moderate levels inhibits academic achievement, (3) the inequitable environmental conditions and resources of segregated classrooms and schools, and (4) the lack of highly qualified, experienced, teachers in these particular classrooms depress learning.

The first source of inequity stems directly from the segregation itself --English learners are more likely to attend classes and schools surrounded by other students who are not proficient in English. This hurts English learners' ability to become proficient in English because research has shown that the composition (relative numbers of English-language learners and fluent English speakers) and structure (opportunities for interaction) of the classroom can inhibit meaningful second language acquisition (Hornberger, 1990; Wong Fillmore, 1991). Moreover, a recent California study found that the higher concentrations of English learners in schools, the lower rates of reading development in first grade (Rumberger & Arellano, 2003).

The educational achievement of English learners is also hurt by their segregation because they are less likely than other students to be surrounded by peers who excel in school. As shown in Table 3, classrooms with high concentrations of English learners also have a higher number of students who are below grade level in reading and math than classrooms with low concentrations of English learners. Research has shown that the academic achievement of peers influences students' own academic achievement, in part, because students learn from each other (Epstein & Karweit, 1983; Hanushek, Kain,Markman, & Rivkin,2001; Hoxby, 2001; Mounts & Steinberg, 1995; Hurd, in press). Thus, the concentration of English learners in California's schools and classrooms not only makes it more difficult for them to learn English, it also makes it more difficult for them to achieve academically.

Lack of integration with native English speakers, both at the school-site and the classroom level can therefore prove problematic for ELs as well. English learners are highly segregated among California's schools, and thus isolated from the language models and social capital necessary to ensure success in school (Harklau, 1994; Olsen, 1997). While most schools have some English learners, the vast majority of these students attend a relatively small percentage of public schools. Thus, English learners are much more likely than their English-only peers to attend schools with large concentrations of EL students. As shown in Table 14, while twenty-five (25) percent of all students in California attend elementary schools in which a majority of the students are English learners, more than half of all English learners (55%) are enrolled in such schools. At the middle school level, only 8 percent of the schools have more than 50 percent of the English learners. Very few high schools have such high concentrations of English learners. Nonetheless, almost half of all EL students attend high schools with more than 25 percent English learners. Thus, the distribution of English learners across schools is uneven and these students tend to be clustered in a relatively small percentage of schools.

Table 14. Schools, Students, and English Learners by Concentration of English Learners and School Level, 1999-2000 (Percent Distribution)

Percent English Learners	Elementary			Middle			High		
	Schools	Students	ELs	Schools	Students	ELs	Schools	Students	ELs
0	6	1	0	<1	1	0	8	2	0
1-25	51	48	15	65	62	30	73	76	49
26-50	24	26	30	26	28	44	17	21	46
51-100	19	25	55	8	9	25	1	1	5
Total percent	100	100	100	100	100	100	100	100	100
Total number	5,306	3,124,107	979,854	1,158	1,059,767	232,481	909	1,538,617	237,129

Note: ELs = English learners.
Source: CBEDS and Language Census.

English learners are even more concentrated at the classroom level. In 2000, researchers from the California Class Size Reduction Study surveyed a representative sample of California teachers who taught grades 1-4 (Stecher & Bohrnstedt, 2002). They found that more than three quarters of all teachers had at least some English learners in their classrooms, and almost one- quarter taught in classrooms with more than 50 percent English learners (Table 15). Even more striking, almost two-thirds of English learners enrolled in grades 1-4 attended classrooms in which more than 50 percent of their fellow students were English learners. Thus, while classrooms in grades 1-4 enrolled an average of 6 English learners (see Table 16) in 2000, the distribution of these students across classes was highly uneven.

Table 15. Teachers, Students, and English Learners in Grades 1-4 by Classroom Concentration of English Learners, 2000 (Percent Distribution)

Percent English Learners in the classroom	Teachers	Students	English Learners
0	24	23	0
1-25	38	39	17
26-50	16	16	20
51-100	22	22	63
Total percent	100	100	100

Note: Results are weighted.
Source: 2000 Class Size Reduction Teacher Survey (N=774).

If students were clustered into these classrooms in order to provide core academic instruction in the primary language and mainstreamed for part of the day to receive instruction in English (preferably in highly interactive and non-high stakes settings like arts, music, physical education), the segregation of EL students would not only be defensible, but would constitute a valid educational treatment. However, in the wake of Proposition 227, most English learners are simply segregated into classrooms populated disproportionately by other English learners where the opportunity to learn both English and academic content is compromised by the lack of appropriate models and instruction targeted to their linguistic strengths.

Table 16. Average number of Students with Selected Characteristics in Classrooms Grades 1-4 by Classroom Concentration of English Learners, 1999-2000

Percent English Learners in the classroom	Total students	Black and Hispanic	English Learners	Free or reduced lunch	Below grade level in reading	Below grade level in math
0	20	4	0	6	4	3
1-25	22	9	3	9	5	4
26-50	21	12	8	14	7	5
51-100	21	17	17	19	9	7
Total	21	10	6	11	6	5

Note: Results are weighted.
Source: 2000 Class Size Reduction Teacher Survey (N=774).

What Can and Should be Done to Address these Inequities?

To some extent all problems of inequity become problems of resources. Disadvantaged students need more resources to help them close the achievement gap vis-à-vis their English only peers. Because redistributing resources is never politically popular, the only solution is almost always to find new resources for those who are shortchanged. While this is easily suggested, it is more difficult to implement. The states, and California in particular, are facing historic budget shortfalls. Realistically, there will be no new state funds for public schools until later in this decade. In the meantime, what can and should be done to address these inequities. We argue that there are things that can be done in every category of inequity.

Increasing access to qualified teachers

The problem of under-qualified teachers is much larger than simply the failure to recruit and train sufficient number of teachers. Close to half of all new teachers recruited into urban districts such as those that disproportionately serve English learners leave the teaching profession within five years (NEA, 2003; Darling-Hammond, 2002). And in schools with high teacher turnover, there is typically high administrator turnover as well. Stability in both leadership and teaching staff is an enormous problem for urban schools. And yet, if there is one area in which district administrators and policymakers can have a significant impact it is this. Administrators must do everything possible to retain well-functioning principals at the same schools, and policymakers must design regulations that help to ensure consistency and stability. We know that teaching and learning conditions are ultimately more important to school personnel than modest salary increases. Every teacher and principal deserves to know they will be backed up by district administration as long as they are doing a good job. Policymakers must find incentives to keep good teachers and principals in their schools, and remove the incentives for them to move into less challenging schools. Of course, this means that administrators and policymakers must tackle, head on, the conditions that drive good people out of these schools—poor facilities, safety concerns, lack of professional support. Administrators must help teachers to create supportive communities, and break down the isolation that teachers often feel, especially in challenging schools. Some schools do this by reorganizing their schedules to allow teachers time, during the school day, to meet, plan, and discuss effective strategies and interventions for students who are struggling in school.

Providing adequate materials and good schooling conditions

Parents are a resource that is too often overlooked for English learners. Even parents without high levels of education can make significant contributions to their children's schools. We know many of the reasons that parents of English learners do not participate actively in their children's schools or schooling, but because parent involvement is known to be an important predictor of students' success in school, it is critically important that the schools find ways to involve them. (Delgado-Gaitán, 1990; Henderson, 1997; Okagaki, Frensch, & Gordan, 1995; Steinberg, Brown, Cider, Kaczmarek, & Lazzaro, 1988; Steinberg, Dornbusch, & Brown, 1992; Steinberg, Brown, & Dornbusch, 1996; Useem, 1992). One important way is to encourage parents to read to their children. Schools must provide books and materials to facilitate this. Most schools that serve large numbers of English learners have –or had—primary language books that can be sent home with children. We know that after the passage of Proposition 227 some schools got rid of these books or put them away (Gándara & Maxwell-Jolly, 2000) but these books should be provided for family reading nights at the school, and be sent home with students.

Beyond this, it is simply **not fair to teachers or students** to be asked to teach and learn in an unsafe, unpleasant environment. Administrators must come forward and demand better from their school district and from their legislators. In economically difficult times policymakers need to make difficult decisions about priorities –what priority does a safe, clean, hospitable school have, especially when we know that this will have an impact on retaining experienced teachers?

Providing equitable and meaningful assessment for English learners

There is no research support for using English language tests to assess students who do not speak English. In acknowledgement of this fact, the NCLB Act requires assessment in the native language where possible, during the initial years that an EL student in our schools. Policymakers should be held accountable to develop an assessment system that is responsive to the needs of English learners, and in the meantime, teachers—especially those who speak the language of the students—should be supported in developing good informal assessments. District administrators should never pass up an opportunity to let legislators know that the current system of assessment of English learners is not meeting anyone's needs. State administrators should seek ways to use federal funds to help develop more appropriate instruments for these students.

Appropriate and rigorous coursework and access to counselors

In part because we draw conclusions about many English learners based on tests that are not capable of accurately portraying their skills, many teachers and administrators underestimate the ability and potential of these students. It is critically important that we hold the same high expectations for English learners that we do for other bright and talented students. We know that there is great variation in the skills that English learners bring to the classroom, and this can create enormous challenges for teachers. But, it is important to capitalize on the strengths that these students have, and many excel in particular areas of the curriculum even though their English skills are weak. Administrators, teachers, and counselors must advocate for their English learners and press for them to be included in all that the schools have to offer. In secondary schools, we must find ways to offer classes that are meaningful and lead to postsecondary options. Even in California, with its heavy restrictions on the use of primary language instruction, at the secondary level, many courses can be provided in the native language of the students, and these courses can meet the same high standard of rigor of any college preparatory class. Moreover, a premium should also be placed on recruiting and hiring counselors who can communicate with English learners and their parents.

Providing more instructional time

Administrators must work with teachers to organize their teaching in ways that reduces the "dead time" that English learners spend waiting for specific instruction and find ways to avoid removing students from content instruction in order to receive English language instruction. Administrators also must avoid ever placing English learners in shortened day or year schedules. These students need more exposure to the curriculum and to models of English, not less.

Teachers and administrators must provide a welcome environment for the parents of English learners. Having parents come in to read to students –in a language they can communicate in—is NOT illegal, and does not violate any aspect of law restricting the use of primary language, as this is not considered instruction. But, it DOES make both students and parents feel more at home in the school and it does teach love of learning and appreciation for the power of literacy and the printed word –prerequisites for becoming good readers.

Meanwhile, policymakers must face the reality that unless English learners are provided **more time** for learning than other students, they will never be able to close the achievement gap. While politically difficult, it is in the interest of everyone that these students perform at higher levels, and this can only be accomplished if policies are created that allow them to receive **more time** dedicated to high quality instruction. In California, more than 500 million dollars are invested annually in after school programs. These programs could be linked much more directly to school instruction and serve as a source of enriched language instruction for English learners.

At the state and national policy level, we need a new initiative on behalf of secondary English learners. Little attention is paid to the needs of students at this level, and their plight is actually far more daunting than that of younger students with more time to adapt.

Increasing professional development for teachers of English learners

Districts provide much of the professional development that teachers receive. School districts must place a higher priority on developing the skills of teachers to work with English learners, and policy makers must press for greater focus on the specific needs of English learners in state supported professional development programs. While "infusing" the needs of English learners into all types of professional development activities is a reasonable ideal, the reality is that it often does not occur at all when it is not the central focus of the instruction. Moreover, the specific linguistic needs of English learners are seldom covered in depth when the professional development is designed principally for English speakers. All professional development activities should include a significant, separate module for teaching English learners.

Addressing the effects of segregation of EL students

Segregation is an enduring –and recurring—problem in American society that has not responded completely to any single intervention (Orfield & Eaton, 1996). And, in some areas of California there simply is not enough diversity of students within a reasonable distance to mix them more heterogeneously. But, we must address this as a serious problem that requires **specific interventions**. Schools can provide language rich environments in which students are exposed to good models of English throughout the day and throughout the curriculum. Many forms of media—newspapers, film and video, audiotapes and radio—drama productions, after school language-based programs that involve local high school and college students can all provide opportunities for students to both hear and speak English.

Because segregation also has implications for what students know about navigating the schooling system, teachers should also be encouraged to hold regular conversations with English learners about "how the system works." What are the things that middle class English speakers know about schooling that English learners do not know? Teachers can help students to acquire more "cultural capital" through systematic, directed, instruction. At the high school level, this can take the form of demonstrating to students the importance of taking higher level math courses, and the social and economic advantages of going to college. We have been surprised by the extent to which many low income and English learner students do not know, for example, the relationship between taking higher math and increasing postsecondary opportunities (Gándara, Gutiérrez, & O'Hara, 2001).

Conclusions

Most English learners are immigrants or the children of immigrants. There is mounting evidence that immigrant students, and the children of immigrants are more academically ambitious than native-born students (see, for example, Suárez-Orozco & Suárez-Orozco, 1996; Rumbaut, 1996). This suggests that there is a critical window of opportunity in which to affect these children's academic futures. If we seize the opportunity and apply the resources while they are in the public schools, we may be able to set these young people on a solid upward trajectory. On the other hand, if we allow this opportunity slip by, the evidence suggests that the challenge will be greater in succeeding generations. The moment is a particularly difficult one, however, for the states, and for California in particular. The unfunded mandates of No Child Left Behind (NCLB) legislation, passed at the beginning of the George W. Bush administration, have placed obligations on the states that many contend they cannot meet. Some people argue that the requirement to staff every classroom in California with a "highly qualified" teacher is beyond the capacity of a state in which nearly one of every five teachers is currently under qualified and there are insufficient resources to meet the needs of a burgeoning student population. Moreover, with the state in virtual bankruptcy, (Note 34) it is unclear where the resources can be found for the increasing numbers of underperforming schools that NCLB requires be aided. In this climate, the special needs of English learners, albeit one-fourth of the school population, are easily overlooked. How is the state to provide a highly qualified teacher for every English learner when it lacks the resources to provide a teacher with *any* qualifications in many of the state's classrooms? And, how is it to ensure appropriate instructional materials and facilities for English learners when it lacks the funds to make capital improvements and repairs on existing facilities that fail to meet basic building code requirements?

California finds itself in the position of reaping the legacy of its own failure to act when economic times were good. Even while the state funded billions of dollars in class size reduction, strengthened the professional development that teachers received across the state, and provided large cash incentives for raised test scores, it failed to make a single major improvement in the instruction of its English learners. While it developed a comprehensive plan for charting the progress of its students with multiple tests and assessments, it did not invest in the development of a single valid measure of academic achievement for its English learners. And, as standards for teacher credentials were raised, the standards to teach English learners were being lowered. Now, as they form a larger and larger portion of the school population, and threaten any real academic gains for the state as a whole, the success of California's reform efforts depends on its ability to raise the achievement of its EL student. Yet there is little evidence that the leadership of the state either understands this urgency or is prepared to address it. English learners in California, and in the nation, represent a potentially rich social and economic resource—if the state invests in them. Without such investment, the future of California education looks grim.

Notes

- 1. National language statistics can be found at http://www.ncela.gwu.edu/states/index.htm
- 2. California population totals can be found at http://data1.cde.ca.gov/dataquest
- 3. In 1999, the state augmented the SAT9 with a test more closely aligned with the state's academic content standards. The first tests, in English/Language arts and mathematics, were first

administered 1999. History/social science and science were added in 2001. In 2003, the SAT9 was replaced with another norm-referenced test, the California Achievement Test (see http://star.cde.ca.gov/).

- 4. The State Department of Education provides aggregate test scores on it's website for each year, but the data are only disaggregated by language groups for the years 1999 through 2001.
- 5. Because of migration and mobility, the cohorts are not necessarily composed of the same students each year, which illustrates the need for a longitudinal study of students (see Kaufman, 2002).
- 6. The CAHSEE covers ELA standards "through Grade 10" and Math standards for "grades 6 and 7 and Algebra I" http://www.cde.ca.gov/statetests/cahsee/background/info.html
- 7. Based on analysis of Early Childhood Longitudinal Study (ECLS) Kindergarten Cohort, California sub-sample (N=2826).
- 8. LAUSD divides its Structured English Immersion classes into two types: Model A, which is English only and Model B, which allows some primary language support. Data are more difficult to interpret for Model A because cell sizes are smaller and the authors report a lack of confidence in these small numbers.
- 9. These include the bilingual certificate of competence (BCC or the Bilingual Cross- cultural Specialist Credential, equivalent to the BCLAD) and the Language development specialist certificate (LDS, equivalent to the CLAD).
- 10. See above.
- 11. The survey did not identify teachers who had authorizations acquired through SB1969 or SB395.
- 12. The Chancellor's Office of the California State University reports that 70% of its credential graduates completed either a CLAD or BCLAD credential.
- 13. This survey, conducted in 2002 by the Lou Harris Polling group, included 1,071 California teachers, both randomly and representatively sampled to approximate a profile of all the state's teachers; 27% were male; 84% were White.
- 14. Portions of this section of the article were written by Kenji Hakuta
- 15. Until 2003, the SAT was part of the state accountability system (STAR) which also includes standards-based test items. In 2003 students were given a new, norm-reference test, the CAT-9 (ETS), in addition to the standards- based items.
- 16. Standardized Testing and Reporting (STAR) Spring 2001 STAR Administration: Frequently Asked Questions. http://www.cde.ca.gov/statetests/star/qanda/smar212001.html
- 17. Academic Performance Index Home Page. http://www.cde.ca.gov/psaa/api/
- 18. About STAR 2001. http://star.cde.ca.gov/star2001/help/AboutSTAR.html
- 19. Standard 11.22 of the AERA/APA/NCME Standards for Educational and Psychological Testing, for example, note that "When circumstances require that a test be administered in the same language to all examinees in a linguistically diverse population, the test user should investigate the validity of the score interpretations for test takers believed to have limited proficiency in the language of the test."
- 20. Standards for Educational and Psychological Testing. (1999).American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Washington, DC: American Educational Research Association.
- 21. The United States Department of Education, Office of Educational Research and Improvement has recently commissioned ARC Associates to conduct a study using San Francisco Unified School District data to help answer this question. We would hope that the findings from this study will inform California testing policy.
- 22. Score Explanations. http://star.cde.ca.gov/star2001/help/AboutScores.html
- 23. San Diego Education Association, CTA Reports, November, 1999
- 24. In a review of test score data for 16 school districts in the aftermath of the passage of Proposition 227, Gándara and Maxwell-Jolly (2000) found few districts actually adhering to this policy. However, the state has not pressured schools to conform to policy and provides no sanctions for failing to do so.
- 25. http://www.cde.ca.gov/psaa/minutes/9905.htm
- 26. www.cde.ca.gov/starpresscharts.pdf
- 27. Results from the May- October 2000 testing window were reported to school districts as late as April of 2004.
- 28. Results from the May- October 2000 testing window were reported to school districts as late as April of 2004.
- 29. Standardized Testing and Reporting (STAR) Spring 2001 STAR Administration: Frequently Asked Questions. http://www.cde.ca.gov/statetests/star/qanda/smar212001.html
- 30. 2002 Stanford 9, Form T, and California Standards Tests. Number of Test Items and Testing Time at Each Grade Level. http://www.cde.ca.gov/statetests/star/2002/staritemstimesSAT 9.pdf
- 31. School districts manage to stay within the law by adding a few minutes at the end of each day to total the same number of hours as students who are on 180 day schedules.
- 32. It is interesting to note that 19 percent of all principals in California reported that their classrooms were never or often not adequate, compared to 9 percent of principals in the rest of the United States.
- 33. This question did not require the respondent to specify in what way the classroom was inadequate.
- 34. At this writing, California was without a budget because no solution could be agreed upon in the legislature as to how to close a gaping $38 billion gap between expenses and revenues, and the state's credit rating was the worst in the country. It was inconceivable that a solution could be found

that did not implicate drastic cuts to social and educational services.

References

Anstrom, K. (1997). Academic achievement for secondary language minority students: Standards, measures, and promising practices. Washington D.C., National Clearinghouse for Bilingual Education.

Artiles, A. J. & Rueda, R. (In press). Factors associated with English learner representation in special education: Emerging evidence from Urban School Districts in California. Cambridge, MA: The Civil Rights Project, Harvard University.

August, D. & Hakuta, K. (1997). *Educating Language Minority Children.*Washington D. C.: The National Research Council, Institute of Medicine.

Berliner, D. C. (1990). What's all the fuss about instructional time? In M. Ben-Peretz and R. Bromme (Eds.), *The Nature of Time in Schools* (pp. 3-35). New York: Teachers College Press.

Brisk, M. (1998). *Bilingual Education: From compensatory to quality schooling.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Brown v. Board of Education, 347 US 483 (1954) (USSC+). 347 US 483.

California Commission on Teacher Credentialing. (2001). Standards of quality and effectiveness for multiple and single subject credentials: Handbook for teacher educators and accreditation team members. Sacramento, CA.

California Department of Education. (2002). Easton releases Spring 2002 California School Exit Exam Results. Press Release. Retrieved January 23, 2003 from http://www.cde.ca.gov/news/releases2002/rel30.asp.

California Department of Education. (2001). Language Census. Retrieved October 1, 2003 from http://www.cde.ca.gov

California Department of Education. (2000). California Standardized Testing and Reporting (STAR), Additional Demographic Reports. Sacramento, California. Retrieved October 1, 2003 from www.cde.ca.gov.

California Department of Education. (1999). Proposition 227 survey: Interim report. Sacramento, CA.

California Education Code Chapter, Sections 60850-60859. High School Exit Examination.

California Tomorrow. (1996). Immigrant Students Project. Oakland, CA.

Callahan, R. (2003). Tracking and English language proficiency: Variable effects on academic achievement of high school ELs. Unpublished dissertation, University of California, Davis.

Carroll, J. B. (1963). A model of school learning. *Teachers college record, 64*(8), 723-733.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F. & York, R. (1966). *Equality of Educational Opportunity.* Washington D. C.: United States Government Printing Office.

Cooper, H., Nye, B., Charlton, K., Lindsay, J. J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*(3), 227-268.

Cornell, C. (1995). Reducing failure of LEP students in the mainstream classroom and why it is important. *The Journal of Education Issues of Language Minority Students, 15,* (Winter).

Crain, R. & Strauss, J. (1985). *School Desegregation and Black Occupational Attainments: Results from a Long-term Experiment.* Johns Hopkins University, Center for Social Organization of Schools.

Crawford v. Honig,, No. C-89-0014 (N.D.Cal.).RFP

Currie, J. and Duncan, T. (1995). Does Head Start make a difference? *American Economic Review, 85*(3), 341-364.

Darling-Hammond, L. (1995). The role of teacher expertise and experience in students' opportunity to learn. *Strategies for linking school finance and students' opportunity to learn.* P. Brown. Washington, D.C., National Governors' Association Center for Policy Research.

Darling-Hammond, L. (2002). Access to quality teaching: An analysis of inequality in California' public schools. Expert report prepared for Williams v. State of California. Retrieved March 8, 2003 from http://www.mofo.com/decentschools/expertreports/darling-hammond report.pdf.

Delgado-Gaitan, C. (1990). *Literacy for empowerment: The role of parents in children's education.* London: Falmer Press.

Diana v. California Board of Education, No. C-70-37. (N.D. Calif. 1970). Complaint for Injunction and Declaratory Relief (Civil Rights), 1973

Durgonoglu, A. (1998). Acquiring literacy in English and Spanish in the United States, in Durgonoglu, A. & Verhoeven, L. (Eds.) *Literacy Development in a Multilingual Context*. Mahwah, N.J.: Lawrence Erlbaum.

Epstein, J. L. & Karweitt, N. (Eds.) (1983). *Friends in School: Patterns of Selection and Influence in Secondary Schools*. New York: Academic Press.

Fleishman, H. L. & Hopstock, P. J. (1993). Descriptive study of services to limited English proficient students. Arlington, VA: Development Associates, Inc.

Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review*, 57(6), 812-28.

Gándara, P. (2000). In the aftermath of the storm: English learners in the post-227 era. *Bilingual Research Journal, 24*(1-2), 1-13.

Gándara, P., Gutierrez, D., & O'Hara, S. (2001). Planning for the future in rural and urban high schools. *Journal of Education for Students Placed at Risk, 6*(1-2), 73-93.

Gándara, P., Maxwell-Jolly, J. et al. (2000). The initial impact of Proposition 227 on the instruction of English Learners. Davis CA, UC Linguistic Minority Research Center Education Policy Center.

Goldenberg, C., Reese, L. *et al.* (1992). Effects of literacy materials from school on Latino children's home experiences and early reading achievement. *American Journal of Education, 100*(4), 497-536.

Goldenberg, C. N. (1994). Promoting early literacy achievement among Spanish-speaking children: Lessons from two studies. In E.H. Hiebert (Ed.) *Getting Reading, Right from the Start: Effective Early Literacy Interventions* (pp. 171-199). Boston: Allyn & Bacon.

Goldman, S. & Rueda, R. (1988). Developing writing skills in bilingual exceptional children. *Exceptional Children, 54*(6), 543-551.

Gottlieb, J., Alter, M., Gottlieb, B. W., & Wishner, J. (1994). Special education in urban America: It's not justifiable for many. *The Journal of Special Education, 27*(4), 453-465.

Graves, A. W., Valles, E.C., & Rueda, R. (2000). Variations in interactive writing instruction: A study in four bilingual special education settings. *Learning Disabilities Research, 15*(1), 1-9.

Gutierrez, K. D., Asato, J., Pacheco, M., Moll, L. C., Olson, K., Horng, E. L., Ruiz, R., Garcia, E., & McCarty, T. L. (2002). "Sounding American": The consequences of new reforms on English language learners. *Reading Research, Quarterly, 37*(3), 328-343.

Gutierrez, K. D., Baquedano-Lopez, P., Alvarez, H. H., & Chiu, M. M. (1999). Building a culture of collaboration through hybrid language practices. *Theory Into Practice, 38*(2), 87-93.

Guy, B., Hasazi, S. B., & Johnson, D. R. (1999). Transition from school to adult life. In M. J. Coutinho and A. C. Repp (Eds.), *Inclusion: The Integration of Students with Disabilities*. Belmont, CA: Wadsworth.

Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* Santa Barbara: University of California Linguistic Minority Research Institute. Policy Report 2000-1

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2001). *Does peer ability affect student achievement?* NBER Working Paper No. 8502 Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2001). *Why public schools lose teachers.* Working paper 8599. Cambridge, MA; National Bureau of Economic Research.

Harklau, L. (1994). Tracking and linguistic minority students: Consequences of ability grouping for second language learners. *Linguistics and Education, 6*(3), 217-244.

Haycock, K. (1998). Good teaching matters: How well-qualified teachers can close the gap. *Thinking K-16, 3*(2), 1-17.

Hayes, K. and Salazar, J. J. (2001). *Evaluation of the Structured English Immersion Program Final Report: Year 1*. Program Evaluation Branch, Los Angeles City Schools.

Heller, K. A., Holtzman, W. H., & Messick, S. (1982). *Placing, Children in Special Education: A Strategy for Equity.* Washington, D.C.: National Academy Press.

Herman, J. & Aguirre-Munoz, Z. (2003). *Preparing teachers to teach English Learners in California: CPDI year 2 evaluation report.* Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST): University of California, Los Angeles.

Herman, J., Goldschmidt, P., & Swigert, S. (2003). *Evaluation of the primary and secondary English Language Arts California professional development institutes: Survey responses to teacher practices and institute evaluation* (01-ELAPDIEVAL-4 and 01-MDPIEVAL-3). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST): University of California, Los Angeles.

Hess, R. D. & Holloway, S. (1984). Family and school as educational institutions. In Parke, R.D., (Ed) *Review of Child Development*

Research (pp. 179-222). Chicago: University of Chicago Press.

Heubert, J.P. & Hauser, R.M. (1999). *High Stakes: Testing for Tracking, Promotion, and Graduation.* Washington, D. C.: National Academy Press.

Hornberger, N. H. (1990). Creating successful learning contexts for bilingual literacy. *Teachers College Record, 9*(2), 212-229.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No. 7867. Cambridge, MA. National Bureau of Economic Research.

Hurd, C. (in press) "Acting out" and being a schoolboy: Performance in an ELD classroom. In Gibson, M. P. Gándara, P. & Koyama, J. (Eds.) *Peers, Schools, and the Achievement of U.S. Mexican Youth.* New York: Teachers College Press.

Hurtado, A., Figueroa, R., *et al.* (Eds.) (1996). *Strategic interventions in education: Expanding the Latina/Latino pipeline.* Latino eligibility study. Santa Cruz, CA: Regents of the University of California.

Jencks. C., Smith, M., Acland, H., Bane, J. J., Cohen, D., Gintis, H., Heynes, B., & Mickelson, R. (1972). *Inequality.* New York: Harper & Row.

Katz, S. R. (1999). Teaching in tensions: Latino immigrant youth, their teachers and the structures of schooling. *Teachers College Record, 100*(4), 809-840.

Kinder, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services: 2000-2001 Summary Report.* Washington, D.C.: U.S. Department of Education, Office of English Language Acquisition, Language Enhancement and Academic Achievement for Limited English Proficient Students. Retrieved October 1, 2003 from http://www.ncela.gwu.edu/states/reports/seareports/0001/sea0001.pdf.

Loeb, S. & Page, M. E. (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *The Review of Economics and Statistics, 82*(3), 393-408.

Lucas, T., Henze, R., & Donato, R. (1990). Promoting the success of Latino language minority students: An exploratory study of six high schools. *Harvard Educational Review, 60*(3), 315-340.

META, Inc. (1996). Comite de padres de Familia, et al. v. the State Superintendent of Public Instruction, et al. San Francisco, CA.

Mickelson, R. A., & Heath, D. (1999). The effects of segregation and tracking on African American high school seniors' academic achievement. *Journal of Negro Education, 68*(4), 566-586.

Milk, R. D. (1990). Preparing ESL and bilingual teachers for changing roles: Immersion for teachers of LEP children. *TESOL Quarterly, 24*(3), 407-427.

Mounts, N. S. & Steinberg, L. (1995). An ecological analysis of peer influence on adolescent grade point average and drug use. *Developmental Psychology, 31*(6), 915-922.

National Association of School Psychologists. (2001). *Membership Directory.*

National Center on Educational Outcomes. (2001).*On the road to accountability: Reporting outcomes for students with disabilities.* Minneapolis, MN: University of Minnesota.

National Research Council. (1998). *Preventing Reading Difficulties in Young Children.* Washington, D. C.: National Academy of Sciences Press.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Oakes, J., & Saunders, M. (2002). Access to textbooks, instructional materials, equipment, and technology: Inadequacy and inequality in California's public schools, found in the Williams Watch Series (wws-rr001-1002). Los Angeles, CA: UCLA/IDEA.

Office of the Chancellor (2001). *First systemwide evaluation of teacher education programs in the California State University: Summary report.* Long Beach, CA: California State University.

Okagaki, L., Frensch, P. A., & Gordon, E. W. (1995). Encouraging school achievement in Mexican American children. *Hispanic Journal of Behavioral Sciences, 17*(2), 160-179.

Olsen, L. (1997). *Made in America.* New York. The Free Press.

Olsen, L., Jaramillo, A., *et al.* (1999). *Igniting change for immigrant students: Portraits of three high schools.* Oakland, CA: California Tomorrow.

Orfield, G. and Eaton, S. E., (1996). *Dismantling desegregation: The quiet reversal of Brown v. Board of Education.* New York: New Press. Distributed by W.W. Norton & Company.

Orr, J., Butler, Y. G., Bousquest, M., & Hakuta, K. (2000). What Can We Learn About the Impact of Proposition 227 From Sat-9

Scores? An Analysis of Results From 2000. Palo Alto, CA: Stanford University.

Ortiz, F. I. (2002) *Essential learning Conditions for California Youth: Educational Facilities*. Los Angeles: University of California IDEA Center. Retrieved October 1, 2003 from http://www.ucla-idea.org

Ovando, C. and Collier, V. (1998) *Bilingual and ESL classrooms: teaching in multicultural contexts*, 2nd ed. Boston: McGraw-Hill.

Page, R. N. (1990). Games of chance: The lower-track curriculum in a college-preparatory high school. *Curriculum Inquiry, 20*(3), 249-281.

Parrish, T. B., Linquanti, R., Merickel, A., Quick, H. E., Laird, J., & Esra, P. (2001). *Effects of the Implementation of Proposition 227 on the Education of English Learners, K - 12. Year One Report*. Palo Alto, CA: American Institutes for Research

Parrish, T. B., Linquanti, R., Merickel, A., Quick, H. E., Laird, J., & Esra, P. (2002). Effects of the Implementation of Proposition 227 on the Education of English Learners. K - 12: Year Two Report. Palo Alto. CA: American Institutes for Research. Retrieved October 1, 2003 from http://lmri.ucsb.edu/resdiss/pdf files/062802yr2finalreport.pdf.

Reagan, T. (1997). The case for applied linguistics in teacher education. *Journal of Teacher Education 48*(3), 185-196.

Reece, J. L., Myers, C. L., & Nofsinger, C. 0. (2000). Retention of academic skills over the summer months in alternative and traditional calendar schools. *Journal of Research and Development in Education, 33*(3), 166-174.

Rodriguez v. LAUSD (1992) No C 6 11-3 5 8.

Rueda, R. (1983). Metalinguistic awareness in monolingual and bilingual mildly retarded children. *NABE Journal, 8*(1), 55-68.

Rueda, R. & Smith, D. (1983). Interpersonal tactics and communicative strategies of Anglo-American and Mexican American mildly retarded and normally achieving students. *Applied Research in Mental Retardation, 4*(2), 153-161.

Ruiz-de-Velasco, J. and Fix, M. (2000). *Overlooked and Undeserved: Immigrant Students in U.S. Secondary Schools*. Washington D. C.: The Urban Institute.

Rumbaut, R. (1996). The crucible within: Ethnic identity, self esteem, and segmented assimilation among American immigrants. In A. Portes (Ed.) *The New Second Generation* (pp. 119-170). New York: Russell Sage Foundation.

Rumberger, R. W. & Arellano, B. (2003). *Understanding and Addressing the Latino Achievement Gap in California*. Berkeley, CA: UC Latino Policy Institute.

Rumberger, R. W., & Gándara, P. (2000). The Schooling of English Learners. In Burr, E., Hayward, G., & Kirst, M. (Eds.) *Crucial Issues in California Education* (pp. 23-44). Berkeley, CA: Policy Analysis for California Education.

Sanders, W. L. and Horn, P. (1995). The Tennessee Value-Added assessment system (TVAAS): Mixed model methodology in educational assessment. In Shinkfield, D. (Ed.), *Teacher evaluation: Guide to effective practice* (pp. 337-350). Boston: Kluwer Academic Publishers.

Sanders, W. L. and Rivers, C. (1996). *Cumulative and residual effects of teachers on future student academic achievement* (pp. 1-15). Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Scarcella, R. and Rumberger, R. W. (2000). Academic English Key to Long Term Success in School. *UC Linguistic Minority Research Institute Newsletter, 9*(Summer), 1-2.

Skirtic, T. M. (1991). The special education paradox: Equity as a way to excellence. *Harvard Educational Review, 61*(2), 148-206.

Snyder, T., Hoffman, L., & Geddes, C. (1998). *State Comparisons of Education Statistics: 1969-70 to 1996-97*. NCES 98-018. Washington D. C.: U. S. Department of Education. National Center for Educational Statistics.

Stanton-Salazar, R. (1997). A social capital framework for understanding the socialization of racial minority children and youth. *Harvard Educational Review, 67*(1), 1-40.

Stecher, B. M. & Bohrnstedt, G. W. (Eds.). (2002). *Class size reduction in California: Findings from 1999-00 and 2000-01*. Sacramento, CA: California Department of Education.

Steinberg, L., Brown, B., Cider, M., Kaczmarek, N., & Lazzaro, C. (1988). *Non-instructional influences on high school achievement. The contribution of parents, peers extracurricular activities, and part time work*. Madison, WI: National Center on Effective Secondary Schools University of Wisconsin, Wisconsin Center for Educational Research.

Steinberg, L., Dornbusch, S. M., & Brown, B. B. (1992). Ethnic differences in adolescent achievement: An ecological perspective. *American Psychologist, 47*(6), 723-729.

Steinberg, L. (1996). *Beyond the classroom: Why school reform has failed and what parents need to do*. New York: Simon & Schuster.

adolescents. Stanford, CA: Stanford University Press.

Tomas Rivera Center (1993). *Resolving a crisis in education: Latino teachers for tomorrow's classrooms*. Claremont, CA.

UC LMRI. 2003. EL Facts, Number 3 Retrieved July 17, 2003 from http://lmri.ucsb.edu/resdiss/2/pdf_files/elfacts_number3.pdf.

Useem, E. L. (1992). Middle schools and math groups: Parents' involvement in children's placement. *Sociology of Education, 65*(4), 263-279.

Walberg, H. J. (1993). Productive use of time. In Anderson, L. W. and Walberg, H. J. (Eds.) *Timepiece: Extending ,and Enhancing Learning Time*. Reston, VA: National Association of Secondary School Principals.

Wells, A. S. & Crain, R. (1994). Perpetuation theory and the long-term effects of school desegregation. *Review of Educational Research, 64*(4), 531-555.

West, J., Denton, K., & Germino-Hausken, E. (2000). *America's Kindergartners*. Washington D. C.: National Center for Education Statistics.

Whitaker, J., Rueda, R., & Prieto, A. (1985). Cognitive performance as a function of bilingualism in students with mild mental retardation. *Mental Retardation, 23*(6), 302-307.

Wiley, D. E. & Harnischfeger, A. (1974). Explosion of a myth? Quantity of schooling and exposure to instruction, major educational vehicles. *Educational Researcher, 3*(4), 7-12.

Wise, L., Sipes, D.E., DeMeyer-Harris, C., George, C., Ford, J. P., & Sun, S. (2002). *Independent evaluation of the California High School Exit Examination (CAHSEE): Analysis of the 2001 administration*. Alexandria, VA: Human Resources Research Organization. Retrieved October 1, 2003, from http://www.ca.cde.gov/statetests/cahsee/eval/2001/2001humrro.html

Wong-Fillmore, L. (1991). Second-language learning in children: A model of language learning in social context. In Bialystok, E. (Ed.), *Language Processing in Bilingual Children* (pp. 49-69). New York: Cambridge University Press,

Wong-Fillmore, L. and Snow, C. (2000). *What teachers need to know about language*. Washington D. C.: U. S. Department of Education, Office of Educational Research and Improvement.

Yaden, D., Tam, A., Madrigal, P., Brassell, D., Massa, J., Altamirano, L. S., & Armendariz, J. (2000). Early literacy of inner-city children" The effects of reading and writing interventions in English and Spanish during the preschool years. *The Reading Teacher, 54*(2), 186-189.

About the Authors

Patricia Gándara

Patricia Gándara is Professor of Education at the University of California, Davis where she directs the Education Policy Center for the University of California's Linguistic Minority Research Institute. She is also Co-Director of Policy Analysis for California Education (PACE), a consortium of the University of California (Berkeley and Davis) and Stanford University.

Russell W. Rumberger

Russell W. Rumberger is a Professor in the Gervirtz Graduate School of Education at the University of California, Santa Barbara and Director of the University of California Linguistic Minority Research Institute. He received a Ph.D. in Education and a M.A. in Economics from Stanford University in 1978 and a B.S. in Electrical Engineering from Carnegie-Mellon University in 1971. He serves on the editorial board of four journals: *American Educational Research Journal, Teachers College Record, Economics of Education Review*, and the *Sociology of Education*. He conducts academic and policy research in two areas of education: education and work, and the schooling of disadvantaged students. His research in the area of education and work has focused on the economic payoffs to schooling and on educational requirements of work. His research on at-risk students has focused on several topics: the causes, consequences, and solutions to the problem of school dropouts; the causes and consequences of student mobility; the schooling of English language learners; and the impact of school segregation on student achievement.

Julie Maxwell-Jolly

Julie Maxwell-Jolly received her Ph.D. from the University of California, Davis, specializing in educational policy and English learners. She has worked as a bilingual teacher and researcher in bilingual and multilingual education and is currently Senior Research Associate to PACE (Policy Analysis for California Education).

Rebecca Callahan

Rebecca Callahan recently completed her PhD. At the University of California, Davis with an emphasis in education policy and English learners. She is currently a post-doctoral scholar at the University of California's Linguistic Minority Research Institute at UC Santa Barbara.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State Univeristy

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Gustavo E. Fischman
Arizona State University
fischman@asu.edu

Associate Editor for Portuguese Language
Pabli Gentili
Laboratório de Políticas Públicas
Universidade do Estado do Rio de Janeiro
pablo@lpp-uerj.net

Founding Associate Editor for Spanish Language (1998-2003)
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

PDF file of this article available:

This article has been retrieved 696 times since October 10, 2003

Volume 11 Number 37 **October 10, 2003** **ISSN 1068-2341**

Systems of Practice: How Leaders Use Artifacts to Create Professional Community in Schools

Richard R. Halverson
University of Wisconsin - Madison

Citation: Halverson, R., (2003, October 10). Systems of practice: How leaders use artifacts to create professional community in schools. *Education Policy Analysis Archives, 11*(37).Retrieved [Date] from http://epaa.asu.edu/epaa/v11n37/.

Abstract

This article explores how local school leaders construct the conditions for professional community in their schools. This paper argues that professional community is a special form of social capital that results, in part, from the design and implementation of facilitating structural networks by instructional leaders in schools. The structural aspects of a school community can be conceived as a *system of practice*, that is, a network of structures, tasks and traditions that create and facilitate complex webs of practice in organizations. Systems of practice are composed of networks of *artifacts*, such as policies, programs and procedures, which can be seen as powerful tools used by local leaders to influence local instructional practices. The system of practice framework suggests that leaders use artifacts to establish structures that facilitate the closure of professional networks among teachers, which in turns builds professional community. The leadership practices of an urban elementary school are used to illustrate how professional community has been developed through the selective design and implementation of artifacts in order to reshape the local system of practice.

Professional community is widely recognized as a valuable quality of local school contexts (Lee and Smith 1996; Little 1982; Seashore Louis and Marks 1996; Newmann and Wehlage, 1995). This paper argues that professional community is generated by networks of trust and obligation developed among teachers and school leaders around shared instructional practices in schools. Social capital is the accumulation of social values such as trustworthiness and respect as a result of participation in networks of social interaction, and "resides in the relationships within an organization and between individuals (Driscoll and Kerchner 1988, 387-388). I argue that professional community is a form of social capital that results, in part, from the work of school leaders to design and implement facilitating structural networks among teachers. The research presented here develops both conceptual tools to make relevant leadership practices visible and analytical tools to show how these practices, taken together, build this special form of social capital in schools.

The paper is organized into two main parts: a theoretical framework designed to capture the coherence and evolution of structures that result in professional community, and an illustration of how the framework is used to analyze leadership practices that developed social capital in an urban elementary school with a demonstrated high level of professional community.

The theoretical framework proposed here explores how the structural aspects of a school community can be conceived as a *system of practice*. A system of practice is the complex network of structures, tasks and traditions that create and facilitate practice in organizations. Systems of practice refer to the structural constraints through which leadership, teaching and learning "flow" in a given school context (Ogawa and Bossert 1995). As opposed to teachers, school leaders often introduce and maintain instructional change in schools through indirect means, such as the development and implementation of programs and policies, rather than through direct engagement with students. Here I describe this indirect influence of leaders on the local system of practice through the design and implementation of *artifacts*. The term artifact, borrowed from human-computer interaction research (c.f. Norman 1988; 1993), refers to entities designed to shape and enable organizational practices. When applied to understanding school leadership, artifacts such as policies, programs and procedures can be seen as powerful tools used by local leaders to influence and maintain instructional practices in schools. A local system of practice refers to the network of artifacts, taken together, that both shape the given context of instruction and point toward opportunities for school leaders to alter instructional practices. A system of practice provides a conceptual framework to explain how leaders use, develop and selectively implement artifacts to influence the practices of teaching and learning in schools.

The study that comprises the second part of the paper profiles an urban school rated to have a high measure of professional community, and asks: 1) what are some of the key artifacts that helped to shape the local system of practice? 2) how did these artifacts evolve together, either by design or by coincidence, to shape the system of practice? and 3) how did the system of practice shape the professional community of the school?

After identifying and discussing the development of three key artifacts, I then use Coleman's (1988) concept of the closure of social systems to show how these artifacts, taken together, create the conditions for professional community in the school. While qualitative data often serve to develop new theories (e.g. Strauss and Corbin 1997), the data discussed here serve as an illustrative example of the theoretical framework described above. The analysis of how leaders in a particular school developed, implemented and used artifacts offers an interesting glimpse into how leaders can create systems of practice that generate professional community, and how researchers and school leaders can re-think their efforts to study and create professional community in schools.

Professional Communities

Professional community provides a model for creating the conditions for their teachers to hear, share and experiment with new ideas about practice. There has been considerable research on the character and effects of professional communities in schools (e.g. Louis, Kruse and Bryk 1995; Bryk, Camburn and Louis 1997; Newmann and Wehlage 1995, Youngs and King 2000; Supovitz and Poglinco 2000). This research indicates that characteristics of schools with strong professional communities include:

- a clear sense of shared purpose and collective responsibility for student learning;
- professional inquiry among staff to achieve that purpose, including opportunities for sustained collaboration and reflection on practice;
- deprivatization of teaching practice;
- norms of collegiality among teachers and leaders;
- opportunities for staff to influence school activities and policies.

Strong professional communities in schools that promote collective responsibility for student learning and norms of collegiality among teachers have been associated with higher levels of student achievement (Lee and Smith 1996; Little 1982; Louis, Marks and Kruse 1996; Newmann and Wehlage 1995).

The concept of "professional community" is a member of the larger conceptual family "communities of practice." A community of practice builds and relies upon a shared core of knowledge through mutual engagement, joint enterprise, and a shared repertoire of skills and abilities (Wenger 1998). In a community of practice, members interact, learn and work through participation in complex networks of shared expectations and norms. Communities of practice include structures and roles that induct new members into core practices through legitimate peripheral participation (Lave and Wenger, 1991). These induction and mentoring structures afford the development of a sophisticated social network to parallel the task networks. While communities of practice often rely on informal structures to facilitate practice, over time, these structures can form institutionalized routines and roles that shape the practice of subsequent members. In more complex organizations, separate communities of practice evolve around common task networks, and can isolate certain groups in the organization from others (Wenger, McDermott and Snyder, 2002). In these cases, community members can find it difficult to transcend established institutional boundaries in order to widen their community of practice. When left unattended, schools and school faculties can fall victim to the peril of institutionalized isolation. Departmental and disciplinary boundaries among teachers (McLaughlin and Talbert 1993; Stodolsky and Grossman 1995) and boundaries between administrative and instructional practice (Rowan 1990; Weick 1976), provide significant obstacles for practitioners to establish common communities of practice across schools. Creating communities of practice that reach across the school provides a considerable challenge for many school leaders.

It is important to note that a community takes on its character from the nature of the practice around which it is organized. While many schools have developed a sense of community among the adults, not all communities can be described as professional. The nature of the practice around which the community is formed proves a key distinction, for example, between a school faculty and a professional community. A professional community is shaped around the goals that define teachers as members of a profession dedicated to promoting student learning, as opposed, for example, to communities organized around student discipline or teacher social interaction (Grossman, Wineburg and Woolworth 2001). Professional communities develop internal practices and expectations to coordinate the non-routine nature of teaching practice through self-regulation and the development of information feedback systems to correct the direction of the community (Louis, Kruse and Bryk, 1994; Huberman 1995; Little and Bird 1987; Argyris 1990). In professional communities, teachers have the opportunities to break down the isolation of classroom in collaborative, problem-setting and -solving activities with colleagues (Halverson 2002; Hargreaves 1994; Huberman 1995; Miller, Lord and Dorney 1994; Rosenholtz 1989). These activities could include collaborative curriculum design, instructional evaluation, interdisciplinary teaming and curriculum development, textbook and course material review, or school improvement planning (Bryk, Rollow, and Pinnell, 1996). Networks of such activities help to create and sustain the conditions for strong professional communities in schools.

Although the value of professional community in schools is widely recognized, knowledge about how individual leaders create and sustain professional communities is not as widely understood. Grossman, Wineburg and Woolworth's experience with developing professional community in a high school led them to comment:

> We have little sense of how teachers forge the bonds of community, struggle to maintain them, work through the inevitable conflicts of social relationships, and form structures for social relationships over time. Without such understanding, we have little to guide us as we create community (2000, 6).

We do have some understanding, however, of what leaders do in schools with strong professional communities. Louis, Kruse and Bryk (1995) conclude that the most important task for school leaders is to create meaningful opportunities for teachers across the school to work together on pressing issues of common interest. Other key leader behaviors include being physically present in the school, creating networks of conversation among faculty; making resources available to support individual teacher development; building bridges to networks practice and knowledge outside the local school; and fostering a school community in which instruction is viewed as problematic.

In many cases, these behaviors both lead to and require structural supports for successful results. Making successful leadership practice accessible means, in part, creating representations of practice to be able to go beyond how leaders create structures to get at how these structures "hang together" in practice. If we assume that professional community is an effect of how these practices together shape a school culture, then we are faced with the need to develop both conceptual tools and practical examples that show both how practices support one another and how aspiring leaders can fashion similar systems in their schools. The knowledge garnered needs to integrate what is known about the *what* of professional community with a framework to show *how* a network of practice can be developed to support such practices.

Distributed Leadership, Artifacts and Tasks

Professional communities do not generate spontaneously in schools (c.f. Grossman, Wineburg and Woolworth, 2000). Rather, school-wide professional community emerges through participation in the activities mentioned above. Much of the responsibility for designing and establishing these activities rests with local school leaders. As discussed above, we know something about the kinds of conditions that both result from and promote professional community, but we do not know as much about how leaders establish these practices in existing school contexts. A distributed perspective on leadership helps to identify and understand the practices that establish the conditions of professional community in schools (Spillane, Halverson and Diamond 2001). A distributed perspective defines instructional leadership as the establishment and maintenance of the conditions for improving teaching and learning in schools (Spillane, Halverson and Diamond 2001, 23). The focus for understanding how leadership is distributed through an organization is to focus on the leadership *tasks*. These tasks are distributed across two primary dimensions in schools: the *social distribution* refers to the network of people engaged in leadership tasks, while *situational distribution* refers to how tasks are constrained and afforded by the context within which leaders work.

I suggest that professional community is an outcome of certain configurations of social networks in a school. Leaders influence the development of social networks not only through direct participation, but also indirectly through the formation of task networks shaped by the design and implementation of artifacts. The concept of *artifact* plays a main role in this argument how leaders build the conditions for professional community in schools. As used in research in human-computer interaction, computer science and cognitive psychology, artifacts are entities intentionally designed to interact with, aid or alter the practices of people (c.f. Norman 1988; Simon 1996; Wartofsky 1979). With respect to schools and leadership, artifacts refer to the programs, procedures and policies designed to shape or reform existing practices in the institutional context (Halverson and Zoltners 2001).

This account of artifacts and leadership relies upon a significant history of research on the institutional and professional structures that frame and enable leadership agency. Early research guiding the Ohio State Leadership Scales, for example, contrasted the concept of "initiating structures," such as schedules and procedures with "consideration," or supportiveness and compassion, to describe how leaders guide the work of followers (Halpin & Winer 1957). The range of initiating structures available to leaders is determined, in part, by the institutional and the culture context of work. Institutional theorists suggest that initiating structures are embedded in institutional routines, and come to constitute the background, framing expectations for work in an organization (Rowan and Miskel 1999). Organizational researchers emphasize how such structures both rely on and help to shape culture. Schein (1992) describes how as organizations grow, they rely on cultural artifacts such as architecture, rituals, stories, and formal statements to perpetuate the established organizational culture. Over time, this network of artifacts comes to constraint the range of possible actions for the organization. Leaders interested in reopening organizational possibilities must engage in the process of deconstructing and rebuilding a new set of artifacts to shape organizational practices. With respect to schools, Deal and Peterson (1990; 1994) consider how leaders need to balance multiple roles in order to attend to how the symbolic and technical structures of schools influence the development of culture. Schools rely upon a network of structures, such as pervasive opportunities for professional development and established occasions to celebrate success in learning and in collaboration, to maintain a positive culture (Peterson & Deal 2002).

While each of these perspectives points out the value of how structures influence and are influenced by leaders, the concept of artifact promises to help us understand the agency of individual leaders in developing structures to influence practice in a given direction. I suggest that the structural context of a school is composed of a variety of artifacts that, over time, come to shape organizational practice. One way to categorize artifacts is according to their place of origin. For example, the situation of school leadership is composed of locally designed, received, and inherited artifacts:

- *Locally designed artifacts* are designed by local actors to address emergent acute and chronic concerns in the school. Locally designed artifacts range from meeting agendas to collaborative curriculum design teams, from daily school schedules and attendance procedures to lunchroom policies. Such artifacts aim to shape practice either through developing a repository of appropriate responses to emergent issues, such as artifacts as that act as precedents for anticipated situations (fire drill policies or appropriate use policies for Internet browsing) or by instituting procedures that routinize practice around intended goals (such as standardized, locally designed curriculum across grade levels, or the structure of the daily school schedule). Locally designed artifacts can, over time, come to be recognized as inherited artifacts (see below) through turnover in leadership or faculty/staff composition.
- *Received artifacts* are adopted and implemented by the local school. These artifacts are received from identifiable external sources, such as state and district authorities, teacher unions, textbook and curriculum publishers, or professional development providers. Examples of received artifacts include policies regarding assessment, budgeting and planning artifacts, or textbooks or curricula. Local institutions are not responsible for the design of received artifacts, but are responsible for artifact implementation and maintenance. The implementation of some received artifacts, such as high-stakes achievement tests and budgeting procedures, is mandatory, in other cases, such as many curriculum packages or student records programs, implementation is optional.
- *Inherited artifacts* comprise the institutional context of practice. Inherited artifacts give rise to practices and routines for which the original artifacts, whether received or designed, have long since been effaced. For example, the nine-month school year resulted from a series of long-lost initiatives to structure the school year according to the planting season; the graded classroom resulted from similar programs designed to create access to education at scale in large urban areas. The specific initiatives that sponsored these practices have long been forgotten-what remains are the ways the artifacts have shaped and institutionalized practices. Local leaders may attempt to correct or mitigate the effects of inherited artifacts either through the implementation of received artifacts or the development of locally designed artifacts.

Both leadership and instructional practice are distributed across a network of locally designed, received and inherited artifacts. Together, this network of artifacts coordinates the practices and routines that form the instructional system of the school. A description of this network, however, is insufficient to get at what leaders do to promote professional community (c.f. Peterson, McCarthey and Elmore 1996). Kruse and Louis (1996) warn "while absence of structural supports impedes professional community; the presence of supportive structures are not sufficient to sustain the growth" (13). An example of the limits of a structural account is the issue of common planning time in school schedules. Establishing programs that build common planning time into the daily schedule is a way school leaders can alter an inherited artifact in order to shape instructional practices. Without meaningful tasks, however, planning time is often spent in non-instructional activities or personal projects. In order to understand how school leaders create and sustain professional community, we must go beyond artifact description to accounts how artifact networks can come to shape school communities.

Systems of Practice

A system of practice is a representation of how the local network of artifacts facilitates the flow of instructional practices of the school. The system of practice is moves beyond a mere context for practice to describe the dynamic interplay of artifact and tasks that inform, constrain and constitute local practice. Teachers and school leaders not only work within the constraints of the network of artifacts in their given situation, but they think about the limits and possibilities of their practice in terms of this network. A school or district-mandated standardized textbook series, for example, provides artifacts that help teachers structure their lessons in certain ways, cover certain material, and understand student learning in terms of an established curriculum. Changing the range of available instructional artifacts not only changes the context of learning, but also influences the ways that teachers understand learning in their classrooms.

This interplay between context and constitution requires a more dynamic, systemic perspective on the conditions leaders establish to shape teaching and learning. Research in activity theory (Engeström 1996) provides a dynamic representational model of contexts that constitute practice. Engeström suggests "contexts are neither containers nor situationally created experiential spaces" (67). Rather, Engeström (1987) proposes that contexts are better seen as *activity systems* that tie the actor(s), the outcomes, and mediating artifacts into a unified system of action. Engeström claims that people engage in the tasks of work through participation in local activity systems. Understanding and communicating work practices requires making the essential aspects of the activity system "visible" for reflection and evaluation (Suchman 1995).

In schools, the practice of teachers and students is constituted by their participation in the activity system of teaching and learning. While researchers have paid considerable attention to the nature of the activity system in schools from an instructional perspective (c.f. Ball and Cohen 1996; McLaughlin and Talbert 1993), school leaders stand in a different relation than do teachers to this instructional activity system. Leaders *qua* leaders do not engage in the activity system of teaching and learning as much as they shape and maintain the system. Leaders are actors on, not actors within, the instructional activity system. This does not mean that teachers cannot be leaders, but it does suggest that as leaders, teachers take a different perspective as participants in the activity system of teaching and learning. Thus schools include at least two levels of activity systems – one frames the practices of teaching and learning, the other frames the practices of school leadership (c.f. Weick 1976; 1982). A key aspect of school leadership is the ability to manage the administrative activity system such that leaders can "make room" to shape the instructional activity system in schools. The ability to engage in both systems simultaneously points toward how management and leadership practices might be integrated in promoting instructional improvement (c.f. Cuban 1988; Elmore 2001).

Considering the activity system of teaching and learning from the outside, as it were, requires that leaders consider the instructional system *as a whole* in order to understand how the different features of the system interact. A system of practice is thus a representation of an external perspective on the instructional activity system from the perspective of leaders – a reification of the activity system for the purpose of identifying the key levers for maintenance and manipulation. Systems of practice reflect leader's perspectives on how the structure of traditions, policies, programs, resources and expectations fit together to shape a school culture and local practices. While the common inherited artifacts of schools create a high level of isomorphism among local systems of practice in ways that provide common constraints and affordances between systems, variations in received and designed artifacts allow local systems of practice to reflect local circumstances distinct for each school. The variation in local systems of practice may explain why artifacts developed and implemented successfully in one setting may be co-opted or marginalized when implanted in another (Powell, Farrar and Cohen 1985; Cuban 1986, 1990). From the perspective of leaders, understanding and learning to manipulate the underlying artifact structure points to areas which can be adjusted to change the tasks of the system in order to support innovative programs. A large measure of local leadership expertise requires getting to know how the unique features of each context influence artifact design and use and understanding how to introduce and manage artifacts that will produce intended changes (Halverson 2002).

Professional Community and the Development of Social Capital

Professional community is an outcome of certain systems of practice in schools. It is evidenced by the emergence of a social network of practice organized around sharing and developing instructional expertise and practice. One way to understand professional community as a form of capacity is to treat it as a special kind of social capital. Capital is used in contemporary economic and sociological discussions to refer to the financial, material or personal resources upon which actors and organizations can draw to maintain or change existing practices. Coleman (1988) developed the concept of social capital to refer to resources available to an actor or an organization by virtue of participation in certain interpersonal or institutional structures. While material and human capital are possessed by the actor personally, social capital "inheres in the structure of relations between actors and among actors" (s98).

Social capital is developed through social interaction (Wehlage 1993). Coleman describes how social capital primarily takes the form of trust among members of a society and an organization. In organizations, trust is built through participation in networks of obligation and commitment, which offer opportunities for participants to rely upon one another in the pursuit of common interests or for the completion of shared tasks. Participation in these networks of reciprocal obligations and commitment help actors to develop reputation in an organization (Fowler 1999). Thus trust is developed as an actor realizes he can work or share ideas with certain colleagues, while reputation accrues when actors in an organization develop opinions about the trustworthiness of other actors.

Organizations with high levels of social capital have high levels of trustworthiness between members. This establishment of trustworthy organizational practices helps people share ideas and abilities together, giving organizational access to resources that had been previously untapped. (Bryk and Schneider 1996) Bryk and Schneider (2002) suggest that a high level of trust among adults in schools is a critical resource for school leaders engaging in program reform. In their examination of Chicago Public School data from 1990 to 1996, they found that schools with high levels of trust at the beginning of reform efforts have a 1 in 2 chance of improving student achievement scores in math and reading, while schools with low levels of trust instead faced a 1 in 7 chance of making significant gains (Bryk and Schneider 2002). While the cause and effect relationship of trust and change is difficult to trace, this research points toward how trust is used as a critical resource for school leaders in organizational change.

While many schools offer ample opportunities for interaction, not all of these interactions help create professional community. Social capital is not a generic capacity – it takes its character from the nature of the interactions from which it is spawned. For example, schools in which adult interactions focus on solving disciplinary and academic problems with individual students, designing individual education plans for special education students, or around teacher social interaction may create social capital, but not necessarily professional community. Grossman, Wineburg and Woolworth (2000) suggest that when conversations around instruction occur in schools with high levels of social capital, but no significant history of professional community, a sense of "pseudo-community" is created in which actors may interact but do not engage in difficult discussions about instruction. In such schools, there are few structured opportunities for interaction about the quality or the process of instruction, and thus little social capital developed around instruction. In the absence of structural supports, it is left to individual teachers to seek out opportunities to interact around instruction. Some teachers develop close relationships with certain colleagues, or engage in professional networks outside the school (Spillane and Thompson 1997; Huberman 1995). When these conversations are left to individual initiative, the social capital that contributes to professional community may be developed among motivated individuals but may not be distributed across the school.

Professional community, then, is a kind of social capital that emerges in certain systems of practice. To create professional community, school leaders either shape existing artifacts or design new artifacts to create the structures that foster social capital. Artifacts that give teachers opportunities to discuss practice, develop programs, and understand assessment information help to create the kind of trust within the organization that marks professional community. The resulting professional community then becomes a form of capacity to support subsequent instructional practice. The next section of the paper provides a profile of the system of practice in a school with a record of strong professional community to illustrate this hypothesis. To highlight features of how local leaders influenced the system of practice, I consider how three key artifacts were created and implmenented to shape the instructional practices on the school, and then describe how these artifacts together helped shape a system of practice that resulted in a strong professional community.

Adams School

To illustrate the how a system of practice yields strong professional community, I have chosen an urban elementary school with a strong professional community as well as a record of improved student achievement. Adams School (a pseudonym), a preK-8 school in Chicago, has an established record of improved student learning, a deserved reputation as a school with a well-articulated vision and record of instructional leadership and professional community, and a stable leadership team willing to grant access to the artifacts that compose the local system of practice. An external report (Consortium for Chicago School Research 1998) indicated high measures of the component aspects of professional community at Adams, including a shared focus on student learning; peer collaboration among teachers and leaders; public classroom practices; reflective dialogue among teachers; willingness for teachers to engage in innovation; and school-wide support for change. During this time, the school also experienced increases in student test scores.

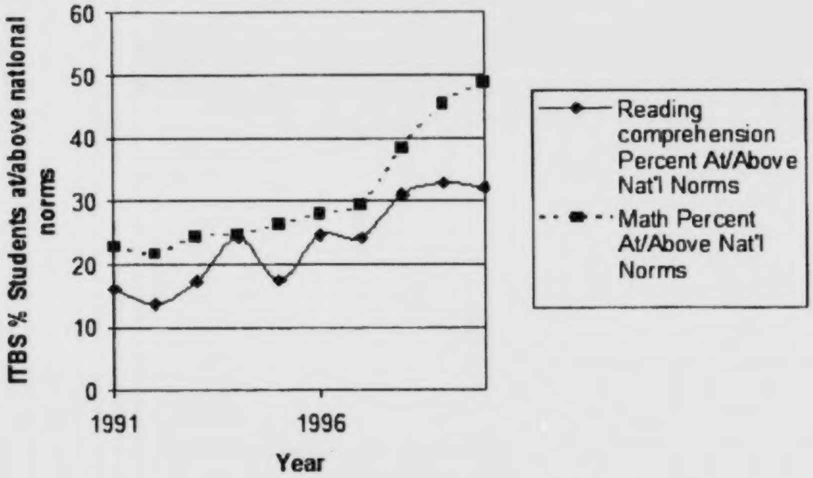


Figure 1. Adams ITBS % students at/above national norms

Measures of student achievement had shown improvement over the period 1995-2001 on the district-wide standardized ITBS (Iowa Tests of Basic Skills) as well as on the statewide assessment IGAP (Illinois Assessment Program). ITBS scores showed significant improvement in student performance in math and reading (Figure 1). These improvements have occurred in the face of annual student mobility rates of 30-40% and the challenge of 97% low-income student population.

The Adams school leadership team was centered around Principal Therese Williams (all pseudonyms). During her twelve-year tenure as principal, Williams led Adams from one of the poorest student performance records in Chicago to a school in which experienced yearly gains in reading and math performance. Williams assembled a leadership team from talented teachers within the building willing to contribute to the creation and implmentation of a series of innovative, locally designed artifacts intended to improve student learning.

The artifacts described here guide the story of how Williams and the Adams school community created professional community and improved student learning. The research presented in here resulted from the collaboration of several research teams to assemble a profile of instructional leadership at Adams.Project researchers made 1-2 visits per week over three years (1998-2000) to record a wide variety of leadership practices. Data collected and developed included multiple structured and semi-structured interviews with leaders and teachers; extensive field notes reporting school meetings and classroom observations; a twenty-three hour video-record of interviews, meeting and classroom observations, and reflective interviews using video as an occasion for discussion; and an extensive catalog of artifacts including school improvement planning documents, teacher observations, meeting agendas, program descriptions, school calendars and schedules, and memoranda.

To access and analyze how leaders used artifacts to shape the system of practice in the school, I looked for evidence of significant artifact use and development, and used the found artifacts found as occasions to analyze the instructional leadership practice in the school. First, the data were coded to identify artifacts either mentioned or apparent in the operation of the school in order to develop a map of the artifacts relevant to instructional practice at Adams school. Eight locally designed or implemented artifacts were identified as components of the local system of practice (See Appendix 1). Once identified, the data were re-considered to understand how the artifacts came to shape the local system of practice. The data were coded a second time in terms of a Design Cycle Analysis Model (DCAM), an analytic model developed to track the genesis, development, iteration and subsequent institutionalization of artifacts (Halverson 2002). DCAM (Appendix 2) was constructed to trace the development of artifacts as outcomes of leader's problem-setting and problem-solving practices. The model seeks to understand how artifacts that result from a problem-setting and solving cycle can come to serve as resources for subsequent problem-setting and artifact design. Conversations with the designers, analysis of the documentary record of artifact development and observations of artifact use were used to explore the component aspects of the DCAM model: the *goals* of the designers, the *strategies* used in the design and implementation of the artifact, the *resources* drawn upon in design and implementation, the situational *constraints* and *affordances* that effected the implementation and use, and the ways in which artifacts evolved over time to become *resources* for subsequent problem-setting efforts.

For this paper, I chose three artifacts to illustrate how Adams leaders attempted to reshape the local system of practice: the Breakfast Club, the Five-Week Assessment program and the School Improvement Planning process. These three were selected as the artifacts recognized most often, both by the researchers and by Adams practitioners, as key to the Adams system of practice. The narratives that follow result from the DCAM analysis of the three artifacts in order to illustrate the genesis and evolution of several key features of the Adams system of practice as well as to show how the artifacts produce the conditions of professional community in the school.

Breakfast Club

Breakfast Club was designed in 1995 as an opportunity for teachers to discuss research relevant to current instructional initiatives and practices among pre K-3 language arts teachers at Adams. Breakfast Club involved monthly meetings in which a teacher led a discussion before the school day about a piece of research, usually concerning reading or writing instruction, with group of pre K-3 teachers and administrators. During the years 1998-2000, there was an average of eight Breakfast Club meetings per year, with an average of fourteen pre K-3 faculty members in attendance. Principal Williams attended about three-quarters of the Breakfast Club meetings during this time period.

Hard-learned experience about the perils of imposing professional development opportunities from above led the school leadership team to consult with a number of grade-level teachers about initial program design. Reflective interviews with members of the design team revealed the following features to be built into the Breakfast Club design:

- the program should not be mandatory to avoid the stultifying atmosphere of many faculty meetings;
- the substance of the discussions themselves should sell the program — if valued information was exchanged at the meeting, word would get around and people would want to come;
- meetings should take place in the mornings, so that teachers would be fresh and ready to entertain new ideas;
- readings should be kept short, so that teachers would have a greater chance of reading them before coming to the session; and
- teachers should be able to select the readings and lead the discussions.

The administrative team thought that the readings should be aligned with the instructional priorities of the school, particularly in language arts, so that teachers would be reading about issues that they should be practicing in their classrooms. Williams thought that a hot breakfast, paid from her own pocket, would give a clear indication to faculty members to show that she was willing to sacrifice for the program to get off the ground.

While Breakfast Club began as an artifact for teachers to talk about research and practice, it has since evolved into a more complex artifact to support teacher brainstorming, experimentation, and design of curricular initiatives. Sample Breakfast Club topics from the 1998-2000 school years included a review of a multiple methods approach to language arts instruction, a conversation about the value and viability of learning centers in primary classrooms, discussions of the components of an ideal language arts classroom, and presentations on how various components of a new school-wide language arts initiative worked out in teachers classrooms. The conversations and interactions that started during Breakfast Club have become a significant organizing framework for the kinds of activities that characterize the local professional community.

Breakfast Club and professional community

The structures and practices of Breakfast Club helped to create some of characteristics of professional community at Adams, including 1) the establishment of teacher collaboration and curriculum design as a cornerstone of the professional development program, 2) the deprivatization of practice and the cultivation and exploitation of in-house expertise among faculty and staff, and 3) the creation of a sense of both vision and ownership about the instructional program.

First, Breakfast Club was originally designed to supplement the existing professional development program at the school. The design represented both a change in degree and a change in kind for prior professional development at Adams. Many externally designed professional development efforts, intended to bring new ideas into the school, proved too intermittent and variable in quality to provide much long-lasting impact on student achievement scores. Early in her tenure, Principal Williams organized curriculum review teams first within grade level (1990-91), then across grade levels (1992-93) to get teachers talking about the school instructional program. Williams attributed the failure of these design efforts to improve test scores to the fact that teachers were merely reorganizing existing ideas instead of importing new ideas into their classrooms and discussions. Breakfast Club extended this significant history of teacher collaboration through the design of an artifact to support group consideration of new instructional ideas. The evolution of Breakfast Club to support teacher-led curriculum experimentation helped spark a change in kind from prior professional development efforts at Adams. Over time, the Breakfast Club discussions came to reflect a blend of reporting on best practices research and teacher reflection on the problems or possibilities offered by their daily practice. The Breakfast Club paradigm helped to change the way Adams leaders and teacher thought about professional development in the school, and created systemic opportunities for teachers to reflect on their instructional practices in light of new ideas.

Second, the opportunity for teachers to lead and participate in Breakfast Club discussions helped to deprivatize practice and created substantial in-house instructional expertise. While initial meetings provided opportunities for interested teachers to become familiar with and discuss new ideas, in later meetings teachers would report on their efforts to try out these ideas in their classrooms. Creating a loop within the teacher community from discussing, to experimenting, to reporting on their experience with new ideas helped to create a system of reflective practice in the school. This was particularly true of the teachers who initially took leadership roles in the discussion and experimentation with new language arts ideas and techniques. The reflective loop created by the implementation of Breakfast Club encouraged many teachers to discuss instructional practices about language arts instruction openly with one another. Deprivatizing practice also had the effect of allowing teachers and school leaders to recognize and exploit the considerable local instructional expertise in the design of subsequent professional development opportunities. For example, spin-offs artifacts such as Teacher Leader (1998) provided a half-day professional development meeting to allow teachers to conduct workshops about the ideas developed and shared during Breakfast Club, while Teacher Talk (1997) applied the format of Breakfast Club to the middle School faculty meetings. The cultivation of in-house expertise, through Breakfast Club and other initiatives, was an important source of developing internal leadership opportunities for teachers within the school. The Adams school leaders developed artifacts such as Breakfast Club, in part, to provide avenues for leadership and the development of expertise, thus helping to enrich the human capital available for subsequent problem-solving opportunities.

Third, Breakfast Club provided an organizing artifact for developing a shared sense of instructional vision and direction. Instead of imposing a sense of direction on the language arts program, the structures and practices of Breakfast Club allowed for the collaborative consideration and experimentation of alternative programs. As teachers explored and reflected upon alternative practices, they could come to realize how the proposed practices might remedy the shortcomings of the existing instructional program. In 1999, after several years of discussion and experimentation, the teachers and school leaders selected Pat Cunningham's Four Blocks of Literacy (Cunningham et. al. 1998) program for the cornerstone of their new language arts program. Breakfast Club served as a foundation for teachers to come together on the need for and merits of instructional initiatives, and provided a structure to support inquiry and collaborative design. The value of Breakfast Club as a structured forum for reflection on practice was shown in several 2000-2001 meetings, as the school community reflected upon their experiences with the Four Blocks program and came to experiment with several new programs to supplement the existing program. Breakfast Club provided a legitimate, on-going forum to discuss and vet proposed directions, helping to continuously test and revamp the plan for language arts instruction in the school.

The structures established by Breakfast Club helped to create practices that resulted in several of the characteristics of professional community in the school. As it began to shape the local system of practice at Adams, local leaders and teachers tinkered with Breakfast Club itself to support an increased range of collaborative activities and reflection on practice in addition to its original goal of bringing new research ideas to the school faculty. This generative effect of the artifact on the system of practice will be explored in the following sections.

Five-Week Assessment

The Five-Week Assessment program was designed as a means to provide meaningful formative data to teachers and leaders about student progress toward improved performance on the summative district standardized tests. At Adams, the ITBS and, more recently, the ISAT presented a challenge for instructional leadership to reshape the instructional program to aid student performance on the district-mandated tests. As a Chicago public school, Adams teachers and leaders are held accountable for demonstrating student achievement improvement as a measure of school performance. The culture of professional community and collaborative design, resulting in part from innovations such as Breakfast Club, has led Adams school leaders to frame the problem of reshaping the school instructional program in terms of collaborative artifact development.

The Five-Week Assessment case offers insight into how the Adams community adjusted to the demands of standardized testing. Every five weeks, teachers throughout the school conducted a 1-2 hour assessment with their students. A team of teachers and leaders collected and graded the assessments, and consequently discussed the results to plan intervention strategies for under performing classrooms. The team also determined the assessment topics. Each year a schedule of assessments was developed for the upcoming school year. Initially designed to prepare students for the ITBS exam, the assessment program shifted toward testing children for the kinds of narrative, expository and persuasive writing and open-ended questions required by the ISAT.

Five-Week Assessment and professional community

Five-Week Assessment was designed meet an emergent need for assessment information within the existing school system of practice. As one teacher described:

> We realized that the (district) tests themselves didn't give us much information about what

> we could do to improve our scores – mainly because we received the results well after we could do anything about it. We thought that a more frequent assessment program, say every nine weeks, would tell us where the children were.

The Five-Week Assessment began as an effort to retrofit the specific, learning outcome demands of the standardized test, particularly in language arts, to the existing instructional system of the school. Prior collaborative design efforts at Adams suggested that this effort too could be an occasion for collaboration. In 1998, a small group of teachers and school leaders worked to establish developmental benchmarks for student achievement by reverse engineering the ITBS.

The initial implementation of the benchmarks provided information about student achievement, but did not suggest what teachers could do to improve achievement. By 2000, the re-designed Five-Week Assessment became an effective diagnostic tool as teachers and leaders collaboratively used the data, through artifacts such as Breakfast Club and Teacher Leader, to shape the existing instructional program by providing intermittent check-points in the curriculum that teachers could use to check student progress school-wide.

While high-stakes accountability systems can provide an occasion to integrate feedback about program effectiveness into the school system of practice, their introduction can also serve to threaten existing professional community in a school. School leaders who use accountability systems to pit teachers, grade levels and schools against one another can erode trust, and lead to a further insulation of practice. At Adams, school leaders realized that using the results of the test scores at the classroom level could create competition and resentment among teachers, and discourage the formation of professional community. The Language Arts Coordinator commented on the need for grade-level reporting of scores to turn accountability data into a positive force:

> I think ... when the IGAP was first started it did something very interesting that almost forced us to work as a team. ... (Reporting at the classroom level led us to think) this one teacher over here could be a shining star, but if the other two or three were not getting the same kinds of results then that one teacher didn't look good anymore because my score was not enough to pull up the entire grade level. So, if I want my grade level to get a good score then I need to help these other teachers pull up to where I am.

The Five-Week Assessment helped to mitigate the summative effect of standardized test scores by providing intermittent benchmarks to gauge the projected results. Although the results of the Five-Week Assessment did not anticipate the standardized test results at first, over time, as the curriculum became more aligned with the assessments, the Five-Week Assessment proved an effective means to point out teachers who were doing particularly well as well as a warning flag for problem classrooms. For example, the Five-Week Assessment (since expanded to include the subjects tested on the ISAT) revealed that 5th grade students in a particular classroom were falling behind in science. The teacher commented that: "looking at the Five Week Assessment saved our butts because we could focus in on helping the students learn the science content they needed to do well on the test." In this case, teachers worked to enhance the existing language arts program with more science-related readings in order to supplement the existing science program. Here the Five-Week Assessment served as an alarm to bring the resources of the Adams professional community to bear in addressing instructional issues before they emerged as accountability problems.

While professional community can emerge from the expression and sharing of common interests around instruction, the long-term viability of professional community may well depend upon the development of feedback structures to provide information about how collaboratively designed initiatives are working. The Five-Week Assessment introduced a mediating artifact between received district accountability measures and the local system of practice in order to make the adjustment of the instructional program tractable, helping to both deepen the professional community and to bring the resources of the community to bear on emergent instructional issues.

School-Improvement Planning Process

Unlike Breakfast Club or the Five-Week Assessment process, the School Improvement Plan (SIP) was a received artifact established as a mandatory district-wide practice for all Chicago Public Schools in 1989 by the Illinois legislature. In many schools, such district-designed instructional planning processes can serve as mandated hoops through which school leaders must jump, completing forms for the sake of compliance and never consulted until the next round of submission is due. When treated as external interventions, such received artifacts can glance off the school system of practice, leaving core instructional practices untouched. However, savvy leaders use features of artifacts such as the SIP to both satisfy district requirements and to stimulate desired instructional changes in the school.

The district-developed school improvement planning process was an artifact designed to help school leaders coordinate budgetary and instructional priorities with the local school councils (LSCs) and the central office. Adams school leaders took the SIP as an opportunity to extend existing collaborative planning practices. School improvement planning was intertwined with many of the leadership practices at Adams, reaching back to the arrival of Principal Williams at Adams in the late 1980s. She reports that instructional planning was one of her initial tasks at Adams:

> (W)e began school improvement immediately, I believe it was 1988 when the first legislation passed that created school improvement plan, and we started from the beginning having everybody who wanted to be involved, involved.

Instructional planning, for Williams, was a way to get faculty and staff involved in conversations around instruction. By the late 1990s, the district-received School Improvement Plan came to serve as a comprehensive artifact to provide coherence to the school professional development and planning processes. Each fall Williams opened the school year with a review of the student achievement goals as specified in the current School Improvement Plan. During the fall semester, teachers would participate in the in-service programs through artifacts such as Breakfast Club and Teacher Talk, and leaders would access the progress of instructional innovations through the Five-Week Assessment. During the spring semester, the community would revisit the School Improvement Plan goals and outline a new plan during a series of formal meetings. In March, subject-matter specific meetings were called to hammer out program priorities and student achievement goals for the upcoming school year. Thus the final plan submitted in May to satisfy district requirements reflected a profound local adaptation of the school improvement planning process to cultivate the local development of professional community.

The School Improvement Plan and professional community

Collaborative inquiry and design are the keys for how the School Improvement Plan process contributed to professional community at Adams. While the School Improvement Plan was itself the outcome of a collaborative design effort, it also served as an "umbrella" artifact to coordinate specific instructional planning opportunities throughout the year, and as a tool to focus the vision of instructional leadership and practice. The role of the School Improvement Plan as an organizing artifact made it a powerful hub for professional community in the school.

Adams school improvement planning provided an on-going, organizing occasion for collaborative design and assessment of the instructional program rather than an isolated task to be completed and shelved. Comprehensive instructional planning, for Williams, was a way to get faculty and staff involved in conversations around instruction. The School Improvement Plan currently plays a central role in organizing multiple collaborative efforts. As described by one school leader:

> (e)verything is tied into in the SIP somehow, that's what gives it credibility in the school. Early on, when the SIP meetings were poorly attended, people would complain about not having the resources to get good work done, and the administrators would reply that the teachers needed to come to the meetings to plan for the things they wanted. The budget, and the initiatives are all tied in, if you want to participate, you have to come early and stay late (at these meetings).

Adams leaders set the problem of school improvement planning as a global process that addresses the key instructional goals of the school, and how, in turn, the instructional goals of the school are customized to satisfy the requirements of the SIP. This iteration between plan and program, between external and locally designed artifacts, shows the compounding effect of interrelated practices over time. The local emphasis on planning also helps to give focus to a shared instructional vision in the school. The School Improvement Plan clearly states both the instructional goals and outlines the means of their achievement; the annual collaborative development of the School Improvement Plan helps insure that the community at large is involved in both understanding and reviewing the instructional mission of the school.

Professional community in action: a vignette

The School Improvement Plan meetings provide a glimpse into the activity of professional community at Adams. In Chicago, the annual School Improvement Plan is expected to outline how the school will support student achievement gains in math and language arts in the upcoming year. A 2000 math School Improvement Plan meeting illustrated how this collaborative planning process worked. Language arts coordinator Gwen Tracy took the lead by instructing teachers to review the 1999-2000 Math plan. After about five minutes of buzzing conversation, a first-grade teacher began a discussion of the adequacy of the current HBJ textbook series. Tracy later explained that:

> The teachers have to own the meeting process because the SIP depends upon their commitment to the changes we propose…if the teachers don't take charge, the meetings don't work….There were a couple of times during the meeting today where (First Grade Teacher Mrs.) Brown looked over at me (for some help at getting the meeting going).

Tracy related that after many of the early SIP meetings, people would come up to her and let her know programs or resources they wanted but didn't bring up at the meeting.

> At first, the teachers didn't see it this way, then they realized that all of the resources are passed out through the SIP – if they weren't involved in the process, they didn't get any of the resources.

As the math discussion unfolded, the five members of the Math Committee (teachers from grades 1, 3, 5, 6 & 8) acted to coordinate the brainstorming session. One Math committee member noted that "We need to work on the more open-ended, problem-solving aspect of math" in anticipation of the new accountability challenges proposed by the ISAT. An eighth grade Math Committee member added that 'next years' (text) book has a lot of practice with open-ended questions…the middle school lessons will have an open-ended question every day…consistent with the NCTM standards." (NCTE is the National Council of Teachers of Mathematics.) Teacher perceptions seemed to be that the while the ITBS focused more on skills testing, new ISAT would focus more on problem-setting and –solving issues. The math committee recognized that the current instructional program was well tailored to the math problems of the ITBS, but not as well suited to the ISAT.

The meeting served as an opportunity to review previous math SIPs plans with respect to other program initiatives. One teacher commented that the Five-Week Assessment program in math be expanded to provide the information generated by the language arts assessments: "I think we should make the assessments similar to how they are planned for Language Arts, I would like to see us plan for the testing in math the same way." This lack of coordination between math and language arts pointed to how the school had chosen to allocate subject-matter leadership resources. Tracy's role in coordinating the Five-Week Assessment in language arts had no organizational analogue in math — the math exams were developed and conducted by full-time teachers and apparently had not received the same attention and review as the language arts exams. This lack of organizational resources was now being felt as teachers faced the new instructional demands of the ISAT. As one teacher commented: "when you look at last years ISATs, (you can see) what we are doing now (for the 5 week assessments) is not working."

This SIP review and design meeting provides a glimpse into the collaborative planning practices at Adams. The meetings are held to provide faculty with an opportunity to shape the school instructional program. The design meetings rely upon considerable resources in developing problem-solutions. Prior experiences with the Five-Week Assessment program, Breakfast Club and collaborative program design meant that teachers and administrators could focus on program refinement rather than novel redesign; experience with group collaboration practices meant that much of the process could be simply assumed so that participants could focus on how programs can be coordinated into a coherent instructional program rather than on the process of collaboration. As one school leader noted,

> (M)ost of the programs we bring up in the SIP are seeded over lunch and at grade level meetings. For example, we talked about the Four Blocks program a full year before we introduced it into the SIP. (One first-grade) teacher who reads a lot presented the basic ideas of the Four Blocks at a Breakfast Club, and there were several Teacher Leader meetings about the Four Blocks program. I know that the program was discussed at grade level meetings, by the time we talked about putting it into the SIP, everyone was on-board.

The School Improvement Plan itself was a district-designed artifact that afforded certain forms of school-level planning, coordination with student achievement outcomes, and discretion over resource allocation. In the hands of Adams school leaders, the plan became an occasion for collaborative design of the school instructional program, and while these practices were not new to the Adams community, the artifact created a powerful and legitimate opportunity for school leaders to deepen and extend the collaborative practices that already existed in the school.

Professional Community and the Closure of Open Systems

Adams school leaders began with a focus on improving student learning, and created artifacts to help teachers understand and develop programs to help students learn better. The intention for the design of programs such as Breakfast Club, Five-Week Assessment or School Improvement Plan was to improve student learning, not necessarily to create professional community. The value of professional community was initially not clear to Principal Williams. After some time, however, she reported that: "we began to believe in the importance of professional community when we realized that, it wasn't taking classes, but that it was when teachers started talking about their teaching that the scores started improving." Professional community was not created so much in the design and implementation of each artifact as realized in the effects of the artifacts taken together, as a system of practice, over time.

If the value of creating professional community was not clear to Adams school leaders, initially the methods of creating professional community were vague as well. As the artifacts began to shape the system of practice at Adams, the emergent sense of professional community helped to create the conditions that helped to shape subsequent artifacts in the school. In other words, professional community was a by-product of instructional improvement efforts that became, over time, a condition for subsequent artifact development. This next section will outline how each artifact created the social capital of professional community within the school, and discuss how the artifacts together helped to for the backbone of a reformed system of practice at Adams.

Coleman (1988) describes how social capital developed through the closure of social or information structures in organizations. Closure happens when actors have opportunities to interact, create trust and develop reputations around selected practices. Closure involves creating feedback loops for information and social interaction in organization. Social capital is developed in organizations and interactions that present redundant opportunities for closure. Open systems, on the other hand, present little opportunity for closure. In open systems, actors diverge from the source of information or directive without structured opportunities for subsequent reconvergence. Trust around core practices does not develop because actors have little opportunity to enter into relations that create obligations or commitments. Many school instructional systems or practice are open in this fashion (Figure 2). In order to promote professional communities in schools, leaders must create legitimate structures that give rise to the occasions in which teachers can share and reflect upon their hard-won instructional expertise, question their own practices, and accept the suggestions of peers. From Coleman's perspective, these structures need to provide *closure* for open social and information networks in organizations. Closing a system means establishing feedback systems in which actors can receive information about the degree to which obligations have been entered into and fulfilled. The instructional systems of many schools remain open as information is distributed within the school with few formal (or informal) structures provided for actors to close the loop. As a result of many mandates and efforts to change instruction in an open systems, teachers and leaders can become disenchanted with received reform artifacts, and quietly learn to insulate their practices from external intervention.

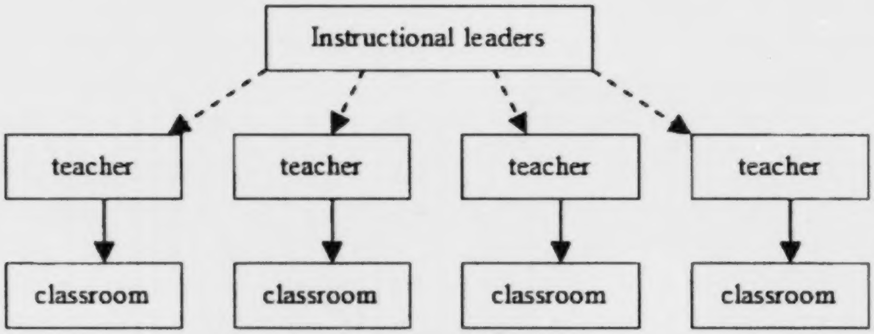


Figure 2. Generic open school

Each of the artifacts described above provides a different form of closure in the local system of practice at Adams. Breakfast Club provides a forum for teachers to reflect both on research and on each other's practice (Figure 3). As it grew to maturity, Breakfast Club added a collaborative design dimension as a platform for the development and customization of the school language arts program. Over time, the communication network among teachers sparked by Breakfast Club became a legitimate venue for developing social capital around instruction among teachers and school leaders, helping to break the barriers among classrooms and with the main office to establish new forms of obligation and trust within the school. Much of the social capital developed during Breakfast Club stemmed from the conscious effort of school leaders to encourage teachers to take leadership roles in conducting and participating in Breakfast Club

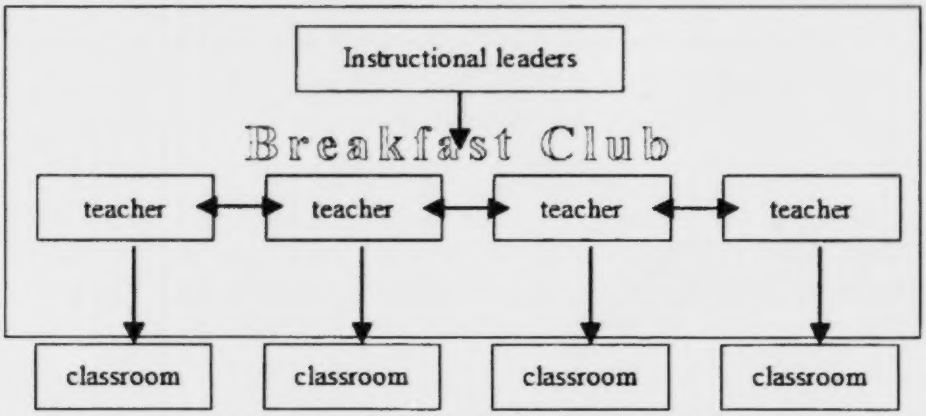


Figure 3. How Breakfast Club closes the system

meetings. The status of Breakfast Club within the school community also helped give the leaders who shape of the discussion agenda and schedule social capital as instructional leaders within the school.

The Five-Week Assessment provides another angle on the on-going effects of classroom practice through collaboratively developed measures of student achievement. Interaction in Breakfast Club was based largely on self-reports of what teachers do in their classrooms. While administrators conduct informal and formal assessments of class="Body" Breakfast Club (Figure 4). The production and discussion of customized quantitative feedback to inform the evaluation of program development helped to create obligations among faculty as teacher look to one another to improve their classroom practice. *De facto* faculty instructional leaders emerged who knew how communicate new ideas with colleagues. The collaborative development and implementation of the Five-Week Assessment provided needed closure among teachers in the system of practice. The Five-Week Assessment also gave school leaders feedback on how new instructional efforts fared in classrooms. Incorporating Five-Week Assessment data into Breakfast Club discussions helped to preserve the tipping point (Gladwell 2000) at which professional community can sustain self-reflective assessment practices without imploding and fragmenting.

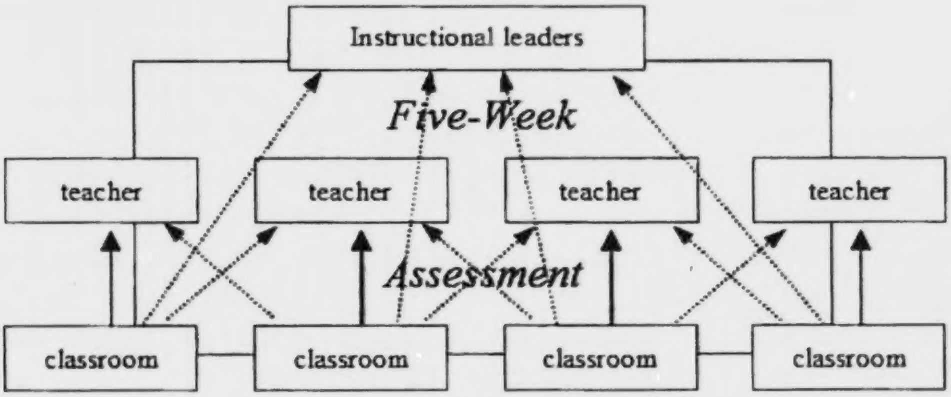

Instructional leaders
Five-Week
teacher
teacher
teacher
teacher
Assessment
classroom
classroom
classroom
classroom

Figure 4. How the Five-Week Assessment closes the system

The School Improvement Planning process augmented social capital developed during Breakfast Club and Five-Week Assessment by allowing teachers and school leaders to articulate not only what they have done, but also to put their ideas to the test by building them into the school-wide instructional program. Since the school was accountable to the district and to the Local School Council (LSC) for achieving the goals specified in the School Improvement Plan, the collaborative planning process gave participants ownership over the direction of the instructional program. The local implementation of the School Improvement Plan at Adams created structures that encouraged multi-level interactions of teachers and leaders in the development of school plans to meet instructional goals (Figure 5).

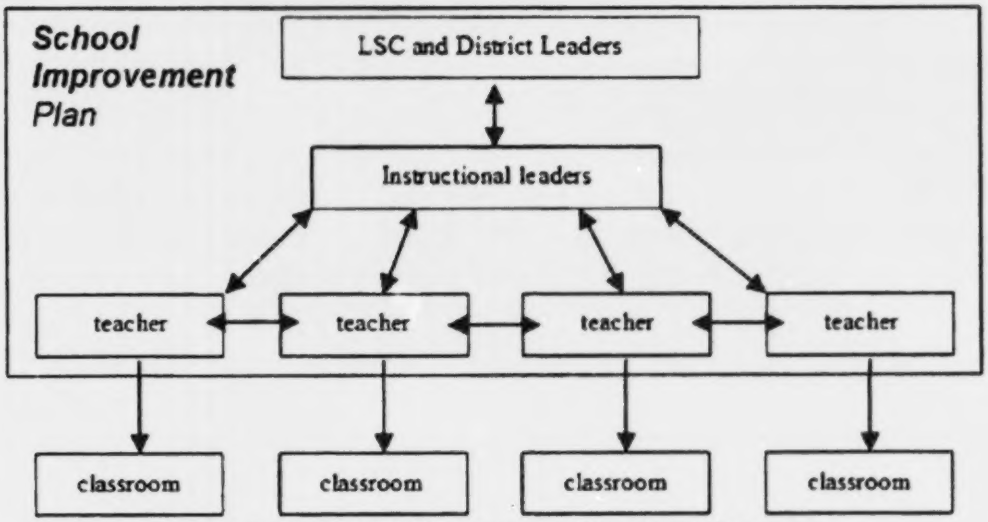

School Improvement Plan
LSC and District Leaders
Instructional leaders
teacher
teacher
teacher
teacher
classroom
classroom
classroom
classroom

Figure 5. How the School Improvement Plan closes the system

While these meetings created obligations among community members to draft and implement viable plans, the successful completion and execution of the plan created trust among members that their work was not in vain.

Separately, the artifacts described here provided structures to support the creation of obligations and trust around instructional issues. Analyzing the function of each artifact in isolation misses the systemic nature of the way the system of practice has evolved at Adams. A school improvement plan, for example, creates neither an atmosphere of innovation nor the means for formative assessment and periodic assessment of practice. Similarly, a five-week assessment that attempts to measure teacher instructional performance progress alone can splinter professional communities because of the threat that comparing teachers to one another make them less likely to collaborate on instructional matters. Together, however, these artifacts helped to create a coherent system of practice that brought closure across the separate artifact-based subsystems (Figure 6). Professional community is the cumulative product of these redundant efforts to close the local system of practice at Adams.

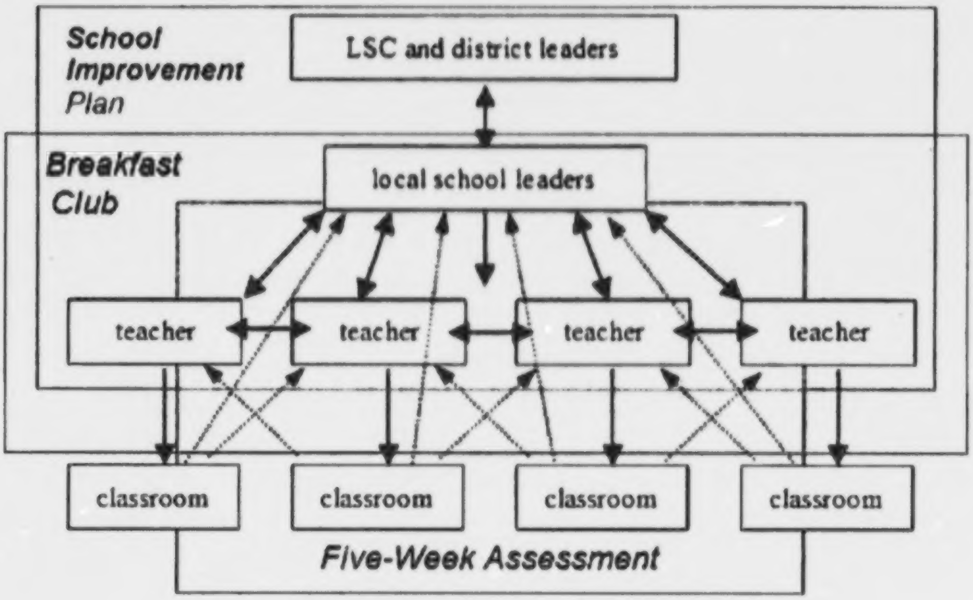


Figure 6. How system of practice closes the system

Discussion

Several interesting issues arise in this analysis about the relation of systems of practice to leadership practice and professional community. First, do artifacts rely on or create professional community? It might be argued either that there was a strong pre-existing sense of professional community at the school upon which these artifact depend for their subsequent success in framing instructional practices at their school. Bryk and Schneider (2002) suggest that existing high levels of trust provide a key resource for school leaders in facilitating school change. Our research showed that there seemed to already have been a pre-existent strong sense of community and shared vision among a tight group of Adams leaders at the school who perceived their responsibility to improve student learning in the school. Perhaps there already also existed a strong sense of professional community among teachers that, when tapped by designed artifacts, blossomed into school-wide professional community.

If professional community can be measured in terms of student learning, however, the effects of the pre-existent professional community were not supported by increases on student test scores. Indeed, in the early 1990s, Adams ranked among the poorest performing schools in the district. One administrator recalled that before Principal Williams, there were strong teachers in the school, and a strong sense of social community among teachers and leaders, but those teachers who initiated discussions about instructional issues felt stigmatized and silenced. While the model provided here cannot conclusively demonstrate causality between artifacts and professional community, it does suggest that the artifacts described above were the key instruments used by school leaders to create trust and open discussions of instructional practice among teachers.

The artifacts themselves, however, do not seem to be easily separable from the context in which they were created. Anecdotal evidence about how other schools that experimented with Breakfast Club-like artifacts felt little impact on the development of professional community suggests that the artifacts themselves are not the answer; rather it is how the artifacts interact with each other and with the existing system of practice to give rise to strong professional communities. Further investigation is required into schools just embarking on the creation of professional community as a avowed outcome to explore the relation between artifact construction and the underlying forms of human and social capital that make professional community possible.

Second, does reliance upon the analysis of artifacts as components of a system of practice give short shrift to the importance of interpersonal and spiritual leadership practices in schools? The analysis of systems of practice offered here is certainly not intended as a comprehensive approach to understanding school leadership practice. Artifacts merely establish the conditions for practice in organizations – the actual practices of teaching and learning involve levels of agency well beyond the determining structures of artifacts. The moral leadership and interpersonal skills required to build consensus, establish vision and give hope in schools transcend the structural components of the instructional context. Still, artifacts provide powerful tools and symbols to convey moral and interpersonal leadership, and the system of practice framework provides a way to understand and access the constraints and affordances that determine what is possible in a given school context. The ability of leaders to create and use artifacts is a powerful capacity not only to shape the practices of teaching and learning but also to provide inspiration through symbolic leadership. The analysis of the artifacts that compose the system of practice by itself may not tell the whole story of instructional leadership, but it does point to a valuable place to start making successful leadership practices accessible to interested others.

Conclusion

This account of how a system of designed and implemented artifacts helped to create a vibrant professional community at Adams provides a vantage point for understanding the nature of professional community as a form of social capital in schools. Looking at systems of practice and the tasks they shape is an important way to consider questions of structure and leadership agency in local schools. Here I have identified a school with a strong sense of professional community, and have attempted to identify contributing artifacts that leaders have used to generate and shape the system of practice in the school. These artifacts taken together help to enable tasks which create and sustain intentional interpersonal relations in schools. School leaders created professional community by using artifacts to shape the local system of practice – creating simultaneous instances of levels of closure that consequently help to form a special kind of social capital. Instructional leadership practice is in part constituted by the ways leaders seek to develop and manipulate the artifacts available within the system of practice. Mapping the artifacts that local leaders create and adapted to shape instruction is an important way to understand the development of professional community. Communicating what these artifacts are and the ways they fit together in practice offers insight of the kinds of situational constructs local leaders build and rely upon in developing local professional communities in their schools.

Acknowledgement

Work on this paper was supported in part by the National Science Foundation and the Spencer Foundation for grants to James P. Spillane for the Distributed Leadership Project, and in part by the National Science Foundation grant to Louis M. Gomez, Daniel C. Edelson and James P. Spillane for the Living Curriculum Project. Additional support was received from a dissertation year fellowship from the Spencer Foundation, from the School of Education and Social Policy at Northwestern University and from the Educational Administration Department at the University of Wisconsin-Madison. The opinions expressed are those of the author and do not necessarily reflect the views of the supporting agencies and institutions. The author wishes to thank Erica Rosenfeld, Louis Gomez, James Spillane, Carolyn Kelley, Allan Odden and Ronald Halverson for their thoughtful comments, and R. Michelle Greene for her research assistance. The author is grateful to the school leaders and teachers who participated in the research.

References

Argyris, C. (1990). *Overcoming organizational defenses: Facilitating organizational learning*. Boston: Allyn and Bacon.

Ball, D. & Cohen, D. K. (1996). Reform by the book: What is — or might be — the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 25*(9): 6-8.

Bryk, A. S., Sebring, P.B., Kerbow, D., Rollow, S., & Easton, J.Q. (1996). Catalyzing basic organizational change at the building level. In *Charting Chicago School Reform*,: Westview Press: Chicago. 93-129

Bryk, A. S., Camburn, E. & Louis, K. S. (1997). Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. Madison, WI: Center on Organization and Restructuring of Schools.

Bryk, A. S. & Schneider, B. (1996). Social trust: A moral resource for school improvement. Chicago: Consortium on Chicago School Research.

Bryk, A. S. & Schneider, B. (1996). *Trust in schools: A core resource for improvement*. New York: Russell Sage.

Bryk, A. S., Rollow, S. G., & Pinnel, G. S. (1996). Urban school development: Literacy as a lever for change. *Educational Policy, 10*

(2), 172-201.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, 95-120.

Consortium for Chicago School Research (1998). *Improving Chicago's schools: (Adams) School.* Chicago: Consortium for Chicago School Research.

Cuban, L. (1988). *The managerial imperative and the practice of leadership in schools.* Albany, NY: State University of New York Press.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920.* New York: Teachers College Press.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19*(1): 3-13.

Cunningham, P. M., Hall, D. P. & Defee, M. (1998). Nonability grouped, multilevel instruction: Eight Years Later. *Reading Teacher, 51.*

Deal, T. E., and Peterson, K. D. (1990) *The principal's role in shaping school culture.* Washington, D.C.: Office of Educational Research and Improvement.

Deal, T. E. & Peterson, K. D. (1994). *The leadership paradox: Balancing logic and artistry in schools.* San Francisco: Jossey-Bass.

DiMaggio. P. J. & Powell, W. W. (Eds.) (1991). *The new institutionalism in organizational analysis.* Chicago, University of Chicago Press.

Driscoll, M. & Kerchner, C. (1988). The implications of social capital for schools, communities, and cities: Educational administration as if a sense of place mattered. In N. Boyan, (1988). *Handbook of research on educational administration.* New York: Longman.

Elmore, R. F. (2000). Building a new structure for school leadership. Washington D.C.: Albert Shanker Institute.

Elmore, R. F. & Fuhrman S. (2001). Holding schools accountable: Is it working? *Phi Delta Kappan, 83*(1): 67-70.

Engeström, Y. (1987). *Learning by expanding.* Helsinki: Orienta-Konsultit.

Engeström, Y. (1996). Developmental studies of work as a testbench of activity theory: The case of primary care medical practice. In S. Chaiklin & J. Lave (Eds.) *Understanding practice: Perspectives on activity and context.* Cambridge: Cambridge University Press.

Fowler, F. C. (1999). Curioser and curioser: New concepts in the rapidly changing landscape of educational administration. *Educational Administration Quarterly, 35*(4), 594-613.

Gladwell, M. (2000). *The tipping point: How little things can make a big difference.* Boston: Little Brown.

Grossman, P., Wineburg, S. & Woolworth, S. (2000). What makes teacher community different from a gathering of teachers? Seattle: Center for the Study of Teaching and Policy.

Halpin A. W., and Winer, B. J. (1957). A factorial study of the leader behavior descriptions. In Stogdill, R. M., and Coons, A. E. (Eds.) *Leader Behavior: Its Description and Measurement.* Columbus, OH: Bureau of Business Research of Ohio State University

Hargreaves, A. (1994). *Changing teachers, changing times.* Toronto: University of Toronto Press.

Halverson, R. (2002). *Representing phronesis: Supporting instructional leadership practice in schools.* Unpublished dissertation. Evanston, IL: Northwestern University.

Halverson, R. & Zoltners, J. (2001). Distribution across artifacts: How designed artifacts illustrate school leadership. Paper presented at the American Educational Research Association Conference, Seattle: WA.

Huberman, M. (1995). Networks that alter teaching: Conceptualizations, exchanges and experiments. *Teachers and Teaching: Theory and Practice* 1, 2.

Kruse, S. D. & Louis, K. S. (1995). An emerging framework for analyzing school-based professional community. In *Professionalism and community: Perspectives on reforming urban schools.* Newbury Park, CA: Corwin Press.

Lave, J. and Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press.

Lee, V. E., & Smith, J. B. (1996). Collective responsibility for learning and its effects on gains in achievement for early secondary school students. *American Journal of Education, 104*(2), 103-147.

Little, J. W. (1982). Norms of collegiality and experimentation. *American Educational Research Journal, 19*(3): 325-340.

Little, J. W. & Bird, T. (1987). Instructional leadership 'close to the classroom' in secondary schools. In W. Greenfield (Ed.)

Instructional leadership: Concept, issues and controversies. Boston: Allyn & Bacon.

Louis, K. S., Kruse, S. D. & Bryk, A. S. (1995). Professionalism and community: What is it and why is it important in urban schools? In K. S. Louis and S. D. Kruse, (Eds.). *Professionalism and community: Perspectives on reforming urban schools*. Thousand Oaks, CA: Sage Publications.

Louis, K. S. & Kruse, S. D. (1995). *Professionalism and community: Perspectives on reforming urban schools*. Thousand Oaks, CA: SAGE Publications.

Louis, K. S., Marks, H., & Kruse, S. D. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal, 33*(4), 757-98.

McLaughlin, M. W. & J. E. Talbert (1993). Contexts that matter for teaching and learning. Stanford University: Center for Research on the Context of Learning.

Meyer, J. W. & B. Rowan (1983). The structure of educational organizations. In M. Meyer and W. R. Scott (eds.) *Organizational environments: Ritual and rationality*. San Francisco Jossey-Bass.

Miller, B., Lord, B., & Dorney, J. (1994). Staff development for teachers. A study of configurations and costs in four districts. Newton, MA: Education Development Center.

Newmann, F. M. & Wehlage, G. G. (1995). Successful school restructuring: A report to the public and educators. University of Wisconsin, Madison, WI: Center on Organization and Restructuring of Schools.

Norman, D. A. (1988). *Psychology of everyday things*. New York: Basic Books.

Norman, D. A. (1993). *Things that makes us smart*. New York: Addison-Wesley.

Ogawa, R. T. & Bossert, S. T. (1995). Leadership as an organizational quality. *Educational Administration Quarterly, 31*(2), 224-43.

Peterson, K. D. & Deal, T. E. (2002). *Shaping school culture fieldbook*. San Francisco: Jossey-Bass.

Peterson, P., McCarthey, S. J. & Elmore, R. F. (1996). Learning from school restructuring. *American Educational Research Journal, 33*(1), 119-153.

Powell, A., Farrar, E. & Cohen, D. K. (1985). *Shopping Mall High School: Winners and losers in the educational marketplace*. Boston: Houghton-Mifflin.

Rosenholtz, S. J. (1989). Workplace conditions that affect teacher quality and commitment: Implications for teacher induction programs. *Elementary School Journal, 89*(4), 421-39.

Rowan, B. (1990). Commitment and control: Alternative strategies for the organizational design of schools. *Review of Research in Education, 16*, 353-389.

Rowan, B. & Miskel, C. (1999). Institutional theory and the study of educational organizations. In J. Murphy and K. Seashore-Louis (Eds.) *Handbook of Research on Educational Administration*. San Francisco: Jossey-Bass

Schein, E. (1992) *Organizational Culture and Leadership*. San Francisco: Jossey-Bass

Seashore-Louis, K. & Marks, H. (1996). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Simon, H. A. (1986). *The science of the artificial*. Cambridge: MIT Press.

Spillane, J. P. & Thompson, C. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis, 19*(2), 185-203.

Spillane, J. P. Halverson, R & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher, 30*(3), 23-27.

Stodolsky, S. S. and Grossman, P. L. (1995). The impact of subject matter on curricular activity: An analysis of five academic subject. *American Educational Research Journal, 32*(2), 227-249.

Suchman, L. (1995). Making work visible. *Communication of the ACM, 38*(9), 227-239.

Supovitz, J. A. & Poglinco. S. M. (2001). Instructional leadership in a standards-based reform. Philadelphia, Pennsylvania: Consortium for Policy Research in Education.

Strauss, A. and Corbin, J. (1997). *Grounded theory in practice*. Thousand Oaks, CA: Sage Publications.

Wartofsky, M. W. (1979). *Models: Representation and scientific understanding*. Boston: Reidel.

Wehlage, G. G. (1993). Social capital and the rebuilding of communities: Issues in restructuring schools, Madison, WI: Center on Organization and Restructuring of Schools.

Wenger, E. McDermott, R. & Snyder, W. M. (2002). *Cultivating communities of practice: A guide to managing knowledge.* Cambridge: Harvard Business School Press.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21*(1), 1-19.

Weick, K. E. (1982). Administering education in loosely coupled schools. *Phi Delta Kappan, 63*(10), 673-76.

Youngs, P. & Kings, M. B. (2000). Professional development that addresses professional community in urban elementary schools. Madison, WI: Center on Organization and Restructuring of Schools.

About the Author

Richard Halverson
School of Education
University of Wisconsin-Madison
Email: halverson@education.wisc.edu

Richard Halversonis an Assistant Professor in Educational Administration at the University of Wisconsin-Madison. His work aims to bring the research methods and practices of the learning sciences to the world of educational leadership. His research focuses on the ways in which people access, learn and teach sophisticated, situated practices such as school leadership. He builds on-line, multimedia technologies to access and document successful school leadership practice, and to develop ideas that can capture the complexity, expertise and situated nature of leadership practice. His recent research interests involve the representation of leadership practices in developing professional community in schools, understanding how local school leaders make sense of and implement teacher-evaluation systems, and representing the ways in which leaders move beyond inclusion to improve learn for all students in their schools.

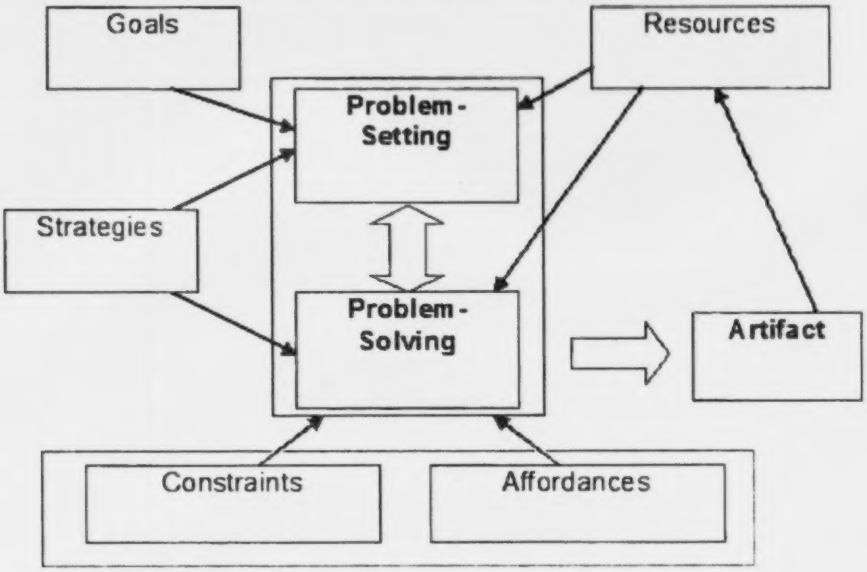
Appendices

Appendix 1: Adams Artifacts

Artifact	Purpose	Description	Designers	Duration of Service
1. Breakfast Club	To provide in-house professional development for and by Adams faculty	Monthly meetings before school at which faculty members make and discuss presentations on research relevant to current instructional programs	Language Arts Coordinator, Principal, Teachers	1995-current
2. School Improvement Plan (SIP)	To create annual local school plan to aligns instructional and budgeting priorities for the upcoming school year.	District designed artifact that acts as a catalyst for local planning efforts as leaders and teachers develop instructional program to meet mandated student test performance targets	District, Principal, Administration, Teachers (approved by Local School Council)	1989-current
3. Five-Week Assessment	Locally-designed testing program to provide formative data to complement summative standardized testing data	Testing program based on reverse engineering summative tests to give teachers and leaders a sense of progress toward improved standardized test achievement	Language Arts Coordinator, Assistant Principal, Principal, Teachers	1995-current
4. Teacher Observation Process	Process to provide formative and summative evaluation of teachers according to union guidelines and district polices	District and locally designed forms used to make sense of principal-teacher observation session. Evaluations based on district and guidelines local instructional program priorities.	District, Principal, Assistant Principal	1989-current
5. Real Men Read	Annual event designed to bring male African American role models into the school to read to the students	An annual breakfast and school wide program in which African-American men gather to eat and read to children throughout the school	Language Arts Coordinator, Assistant Principal, Principal	1998-current
6. Career Day	Annual event designed to offer Adams students an opportunity to survey career possibilities.	A two-part annual assembly for middle school students to listen to African-American speakers, then meet with African-American professionals in a variety of career fields.	Guidance counselor, principal, teachers	1999-current
7. Chicago	Year-long curriculum	Collaborative curriculum design	Science coordinator,	2000-

Annenberg Challenge Curriculum Planning Process (CAC)	planning process to document collaborative design efforts in building multidisciplinary curricula	effort using LeTUS project-based science curricula as a seed for building middle-school cross-disciplinary curriculum	Teachers, Northwestern and Roosevelt University Researchers	2001
8. Science Coordinator Position	Position established to design science program for Adams' designation as Math-Science Academy	Promotion of 6th grade teacher Tim Zacharias to renovate science program and to design and teach middle school science curriculum in collaboration with classroom teachers	Science coordinator, Principal, Assistant Principal	1999-2000

Appendix 2. Design Cycle Analysis Model (Halverson 2002)



The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State Univeristy

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh

University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Gustavo E. Fischman
Arizona State University
fischman@asu.edu

Associate Editor for Portuguese Language
Pabli Gentili
Laboratório de Políticas Públicas
Universidade do Estado do Rio de Janeiro
pablo@lpp-uerj.net

Founding Associate Editor for Spanish Language (1998-2003)
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA is published by the Education Policy Studies Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

PDF file of this article available:

This article has been retrieved 297 times since October 17, 2003

Volume 11 Number 38 **October 17, 2003** **ISSN 1068-2341**

The Allures and Illusions of Modernity: Technology and Educational Reform in Egypt

Mark Warschauer
University of California, Irvine

Citation: Warschauer, M., (2003, October 17). The Allures and Illusions of Modernity: Technology and Educational Reform in Egypt. *Education Policy Analysis Archives, 11*(38).Retrieved [Date] from http://epaa.asu.edu/epaa/v11n38/.

Abstract

Much of the research to date on educational technology has focused on its implementation in wealthy countries. Yet instructional technology has a special allure in the developing world, where it holds the promise not just of improving schools but also of hastening modernization. This article examines a national educational technology effort in Egypt, illuminating the contradictions between the rhetoric of reform and the reality of school practices. The analysis points to underlying political, cultural, and economic factors that constrain attempts to improve Egyptian schooling with technology.

Educational technology has always been about much more than improving learning. In the eyes of many proponents, it has been about *transforming* learning -- overcoming traditional educational approaches and supplanting them with revolutionary new paradigms of teaching, learning, and schooling. The flashy new machine in the classroom—whether the film projector or the television or the computer—has represented the pinnacle of modernity in the eyes of its supporters, whether from the government, the business sector, or academia.

Larry Cuban (1986) has done an excellent job of chronicling America's 100-year love affair with gleaming new machines in the classroom, from the radio to the computer lab. Each era has experienced the same cycle of bold promises followed by erratic and disappointing diffusion, with the technology eventually finding a small niche on the margins of the educational process. Where education has changed in the process, it has been in the elite schools of the well-to-do that were disposed to reform in the first place (Cuban, 1986; 1993b).

Now the cycle begins again, with the introduction of the computer and the Internet. There has certainly been no shortage of bold claims about how computers will revolutionize the classroom, transforming the teacher from the stereotypic *sage on the stage* to the new and equally stereotypic *guide on the side* (Knapp & Glenn, 1996; Means, 1998; Mehlinger, 1996; Sandholtz, Ringstaff, & Dwyer, 1997; Starr, 1996). Learners will become autonomous and goal-directed, classrooms will become centers of collaborative, critical inquiry, and technology will have finally transformed schools to match the needs of the information society (see, for example, Starr, 1996).

Research to date, however, makes such claims questionable. Most studies show that the use of computers tends to amplify whatever prior approaches and processes were already occurring in classrooms, rather than transform them (e.g., Warschauer, 1999; 2000). For example, a recently completed four-year national US study of "network science" – in which learners from throughout the world collect and share scientific data over the Internet – found that the projects tended to trivialize rather than transform learning, unless they were based on teacher-led practices of scientific inquiry in the individual classrooms (Feldman, Konold, & Coulter, 2000).

In fairness, it is too early to judge the lasting impact of computer and Internet technology in the classroom. Many people believe that the computer and Internet have a more direct relationship to fundamental changes in human communication and cognition (see Harnad, 1991) and the overall organization of the economy and society (see Castells, 1996; 1997; 1998) than did previous technologies such as television or film. Thus even those who have taken a hardheaded and realistic look at computers in the classroom, such as Becker (2000; 1982) believe that under the right conditions it may facilitate educational reform (Becker, 2000). Perhaps the best that can be said about this is that the jury is still out.

Not surprisingly, the discussion to date of educational technology has taken a US- and Euro-centric viewpoint. The penetration of information and communication technologies (ICT) in most of the rest of the world is much lower, whether in the office, the shop floor, or the classroom. With only a small elite having computer access, and the majority of their citizens living on a few dollars a day (United Nations Development Programme, 2000), the developing countries in South Asia, Africa, and Latin America have not yet been able to fill their classrooms with computers and Internet connections.

However, the lack of modern technology does much to heighten its allure in the developing world. There is much discussion of the potential of new technologies to help countries leapfrog out of underdevelopment. Just as Germany and Japan—with their infrastructure destroyed after World War II—used a completely new infrastructure to catch up to or even leap ahead of other capitalist countries in efficiency of production, many believe that the least developed countries can now make use of information and communications technologies to skip over stages of development (see discussion in Singh, 1999). And indeed, it is precisely those countries that have been able to make effective use of information technologies, such as Singapore, Korea, and Taiwan, which have most recently progressed from underdevelopment to the ranks of the wealthier countries. For developing nations, information technology thus holds the allure of allowing rapid entry to modernity.

A number of developing and middle-income countries in Latin America, Asia, and Africa are now beginning to experiment with information technology in the classroom. With the price of computers and telecommunications falling, and schooling in many countries badly needing improvement, developing countries have great incentive to try to integrate new technologies and new approaches (Osin, 1998). To date, though, little research or analysis has been published on why or how developing countries are attempting to make use of technology in schooling, or what the results have been. The few exceptions to date have for the most part been descriptions of model projects supported by international donor agencies (e.g., Calderoni, 1998; Potashnik, 1996). Though these reports have been helpful, they tend to focus on best practices rather than shedding light on actual practices. The lack of broader and more in-depth analysis of educational technology practices in developing countries can unfortunately lead to a situation whereby educators in those countries uncritically mimic the practices (or what they may falsely believe to be the practices) of wealthy countries, without proper regard to local conditions and circumstances, thus worsening rather than solving problems. Such counterproductive approaches are heightened by the fears of being left behind in the information revolution (see discussion in Agre, 1997).

To help overcome this lack of information on educational technology in the developing world, I carried out a three-year qualitative study in Egypt. Though, as in any such study, the findings apply in particular to the situation under investigation, Egypt represents an excellent example of a society poised on the edge between underdevelopment and modernity (see, for example, New and Old: A Survey of Egypt, 1999).

The study focused on the policies and practices of integrating technology in education in governmental K-12 schools, under the leadership of the Egyptian Ministry of Education. The overall unit of analysis for this study is the governmental K-12 educational sector in Egypt. Where relevant, I also consider data from other educational units in Egypt, including K-12 private schools, governmental and private universities, and non-governmental community technology centers. Data sources for the study included the following:

1. *Participant observation*: I engaged in participation-observation continuously for three years, from 1998-2001, while I was involved in a donor-funded educational project in Egypt. During this time, I participated in efforts to plan, implement, and evaluate technology-based interventions in Egyptian schools and inservice and pre-service teacher education programs. My participation included attendance at meetings of Ministry of Education bodies, international donor and implementation agencies, and Egyptian non-governmental organizations, as well as attendance at and participation in in-service and preservice teacher training programs. It also included professional visits to 25

Egyptian primary, preparatory (i.e., middle), and secondary schools located in rural and urban areas throughout the country and to colleges of education in 10 Egyptian universities, and participation in meetings and training sessions among Egyptian educators. I took notes during these visits and sessions and afterwards typed them up in personal and professional reports. Finally, I have participated in various electronic discussion forums of Egyptian educators focused on use of technology in education.

2. *Interviews and focus groups*: I conducted approximately 100 individual interviews with Ministry of Education (MOE) officials, business leaders, representatives of non-governmental organizations, parents, and students. I also organized about ten focus group meetings of six-to-ten K-12 teachers and faculty members at colleges of education to discuss the integration of technology in classrooms and programs. I took notes during these interviews and focus group meetings and typed them in personal and professional reports following the interviews and meetings. The interviews and focus groups were organized within the context of my work in Egypt and addressed issues related to access to technology at educational sites, skill and knowledge level of educators, and goals and objectives of using technology with students and in professional development programs.
3. *Analysis of documents*. I have collected and analyzed a wide array of documents and reports issued by the Egyptian government and MOE, donor agencies, and non-governmental bodies. While the majority of these are in print, they also include electronic documents, such as Websites of MOE bodies and schools.

The study draws on critical approaches to research on infusion of information and communication technologies (Warschauer, 1998). A critical theory of technology (see Feenberg, 1991) distinguishes itself from both *determinist* approaches (which view technology as of necessity having a positive or negative impact) and *instrumental* approaches (which view technology as a valueless tool which can be deployed toward any end), and critical approaches. (Determinist approaches are alternatively referred to as substantive approaches or autonomous approaches (see, for example, the work of Ellul, 1980). Instrumental approaches are alternatively referred to as neutralist approaches and are often backed by technologists; see discussion in Shallis (1984).) Both determinist and instrumental approaches are seen as downplaying the embednesses of social, political, economic, and culture factors in technologies, which shape (but do not determine) how technologies are deployed. In a critical approach, technology is viewed as a site of struggle, and investigations of technology implementation seek to uncover underlying power relations that shape how technology is used, similarly, for example, to how critical literacy studies seek to uncover the underlying power relations framing literacy practices (e.g., Street, 1984; 1993).

In reporting on the study, I will focus on three aspects: (1) the discourses of technology-based educational reform, (2) the practices of educational technology, and (3) the social context of education and technology which helps explain the (mis)match between discourse and practice. First, though, I will briefly introduce some necessary background information on Egypt.

Egypt at the Turn of the Millennium

I decided to conduct this study during my first week in Egypt, as I stood on the banks of the Nile and took in the Cairo landscape. Across the Nile, I saw the glimmering towers of the World Trade Center, including some of the fanciest stores, restaurants, and offices of modern Egypt. Looking down, though, I also saw a poor family of eight who lived in three tiny boats by the bank of the Nile. Thin and poorly clothed, this family apparently spent their days and nights on a couple of tiny canoes no longer than a fishing pole. Yet, as I looked down, I saw a shiny object in the center boat, and, upon looking more closely, I realized it was a battery-operated television. Even this impoverished family living in tiny canoes on a highly-polluted river was grasping at modernization through media. This contrast within contrasts was an excellent introduction to me of Egypt today. Egypt is rushing toward modernization, while at the same time modernization must conform itself to the centuries-old ways of life of Egyptian society.

The use of technology in education in Egypt is situated in a broader social and educational reform movement that dates to the early 1990s. In 1992, the Egyptian government, backed by the World Bank and International Monetary Fund-backed structural adjustment program, launched an ambitious structural adjustment program (Korayem, 1997). The ongoing program seeks to transform the previously stagnant, insular semi-socialist economy inherited from the Nasser era into a modern, transparent, and efficient economy that can compete in a global market (Galal, 1995; Sachs, 1996). The reform process has shown some positive results; Egypt's gross national product grew on an average of 5.4% annually from 1995-2000 (based on data available from the World Bank, available at http://devdata.worldbank.org/data-query/), up from an average of 1.5% in 1990-1995 (Galal, 1995). This growth has brought Egypt from the ranks of the least developed countries up to the lower medium-development countries, ranking at 119 out of 174 countries according to the United Nations Development Programme's Human Development Index, with a gross domestic product of $3041 per person (GDP is calculated by the UNDP according to *purchasing power parity*, i.e., how many equivalent goods can actually be purchased in a country), an average life expectancy of 66.7 years, and an infant mortality rate of 6.9% (United Nations Development Programme, 2000).

In spite of some areas of improvement, Egypt is still troubled by high rates of poverty, reflected by a low literacy rate and poor public health in urban and rural areas, and the financial structures of the state are far from being fully reformed or stable (Institute of National Planning, 1998; New and old: A survey of Egypt, 1999). The structural adjustment program thus continues with the goal of modernizing the economy, overcoming social exclusion and poverty, and bring Egypt to the economic level of middle-income countries such as Malaysia and eventually to that of newly-industrialized countries such as Korea.

In no other arena is the need for institutional reform and social inclusion greater than in education. There is wide consensus among both educators (e.g., Jarrar & Massialas, 1992; Tawila, Lloyd, Bensch, & Wassef, 2000) and economists (e.g., Bartsch, 1995; Fergany, 1998) about the poor performance of Egyptian schools, even when compared to that of other developing countries (Birdsall & Lesley, 1999). Problems identified include large class sizes, often exceeding 45 students per class in urban schools; poorly trained teachers with low wages and status; and a centralized, test-driven curriculum focusing on rote memorization of unimportant material (Jarrar & Massialas, 1992; Ministry of Education, 1993; Tawila et al., 2000). These problems are reflected in a $2 billion private-tutoring industry that is half the size of government expenditures on public education (see discussion in Birdsall & Lesley, 1999). In effect, teachers have a disincentive to teach well, since they earn for more than their governmental salary by tutoring their own students privately to make up for what they failed to learn in school.

The expense and low quality of education have contributed to a high dropout rate in primary school (Fergany, Farmaz, & Wissa, 1996) and a corresponding low rate of adult literacy (53.7% overall and only 41.8% of women, United Nations Development Programme, 2000). They also lead to a low-skill level and employment potential among those who complete school; two studies claim that high school graduates in Egypt who don't go on to university have less earning potential than people who have only partially completed primary school (Bartsch, 1995; Fergany, 1998).

The problems of education in Egypt are systemic and stem from a wide variety of causes, including the poor state of education following the era of British colonial influence, the rapid population growth rate which overwhelms limited resources, and the priorities of the previous Nasserist system which emphasized the quantity of schools rather than their quality (Jarrar & Massialas, 1992). Also of note is the limited demand for education, at least in the past, due to the poor-performing Nasserist economy (Birdsall & Lesley, 1999). With economic reform and growth major governmental priorities, there is now widespread recognition in Egypt that poor-performing schools are a drag on socio-economic development and that educational change is critical.

ICT in Egypt

The other major contextual factor shaping technology use in education is the general growth and role of information and communication technologies (ICT) in Egypt today. Egypt began emphasizing the adaptation and integration of ICT in the early 1990s. Expansion of ICT is viewed as critical for modernizing production, distribution, and marketing efforts and thus assisting Egypt in competing successfully in the global market (Mintz, 1999). Egyptian government and business leaders also hope that the information technology sector will become an important industry in its own regard, and they often point to India and Israel as models that they would hope to emulate. Egypt has thus placed major emphasis on ICT, and Egypt is reputed to be one of the fastest growing ICT markets in the world. New communications media in Egypt include the Internet, mobile telephony, and digital satellite television.

The Internet. The Internet was first introduced to Egypt in 1993, when a small university network was established (Information technology in Egypt, 1998). Commercial Internet use began three years later and has developed with more government support and less censorship than in many other Mideast countries, reaching a total of some 600,000 Internet users by 2001 (NUA, 2003), representing about 1% of the population. The growth of the Internet in Egypt is constrained by economic factors in a country where per capita income is roughly $120 per month (World Bank, 2001). This is compounded by the fact that local telephone calls cost $1-$3 an hour, making frequent Internet use expensive even for the small middle class. Low teledensity rates – 6% nationally (United Nations Development Programme, 2000) and only 2% in rural areas (Badawi, 2000, July)– mean that people do not have telephones to log on, and only about 1% of the population own computers (United Nations Development Programme, 2000). The Internet is thus inaccessible to Egypt's poor, and even many in the small middle class must resort to coping mechanisms, such as sharing Internet accounts or using Internet cafés.

Economics is not the only factor restricting access to the Internet. Other major factors are the high illiteracy rate and language usage. The Internet largely arose in Egypt in English-language milieu, including the country's small high-tech and international and foreign business sectors, and to this day common standards of Arabic language computing and communications have not been reached. That means that the vast majority of Web sites and computer-mediated communication is conducted in English (Warschauer, Refaat, & Zohry, 2000). This presents less of a problem for the Egyptian elite, many of who have studied in English medium schools and can thus read and write the language as well as Arabic. However, English is taught very poorly in public schools, so the vast majority of the people do not know it at all.

Wireless telephony. Egypt has also tried to extend its new media through wireless telephony. The number of lines grew to more than 1 million in four years (El-Nawawy, 2000), thus swamping Internet growth. The higher rate of wireless telephone use compared to Internet use is due to a variety of reasons, including language (telephone communication is done in Arabic), initial investment (a wireless phone is much less expensive than a computer), and the familiarity of phone use. Wireless telephony penetration has reportedly tripled again from 2000 to 2002 to reach three million lines (Arab Communication Consult, 2002), or about five percent of the population.

Satellite television. Finally, the government has invested heavily in the development and launching of two digital television satellites, Nilesat 101 and Nilesat 102, with some 180 stations, as another medium of high-tech communications. Nilesat is designed to serve developmental goals through its emphasis on educational program (see discussion below). Nevertheless, with reception requiring not only a television and satellite, but also a $400 digital receiver, Nilesat is believed to have very few subscribers to date (Sakr, 1999).

In summary, then though new information and communications media have grown rapidly, they remain accessible only to a few percent of the Egyptian population. The urban and rural poor in most need of access to information and communication resources are excluded from the new media.

It is not surprising that in Egypt, as elsewhere, information and communication technologies are being used principally by those with money. This reflects the natural amplifying affect of the ICT throughout the world: those with financial, human, and social capital have better access to ICT, which they can use to further enhance their financial, human, and social capital. And indeed, no matter how well motivated the Egyptian government or private sector were, there is no way they could instantaneously put computer, Internet connections, mobile telephones, and satellite televisions in the homes of Egyptians poor.

Nevertheless, at an institutional level, governments can deploy ICT to serve broader developmental goals. In Egypt, the main sector in which Egypt has attempted to deploy ICT for broader development purposes is education. The discourses of technology-based educational reform, and their practices, will now be discussed.

Discourses of Technology-Based Reform

The Government of Egypt believes that it has found a perfect combination in technology and educational reform. It has an ambitious and expensive plan to use ICT to help overcome the country's educational problems while simultaneously preparing a technologically-skilled workforce to meet the demands of the 21st century.

The Ministry of Education (MOE) initiated its national plan for the technological development of education in 1994. A special unit within the MOE, called the Technology Development Center (TDC) was formed shortly thereafter to coordinate the MOE's effort to infuse technology into schools.

The goals of the national technology in education plan have been laid out in a number of publications issued by in the name of the TDC (e.g., Technology Development Center, 1997), the MOE (e.g., Ministry of Education, 1999), and the Minister of Education, Dr. Hussein Kamel Bahaa El Din (e.g., Bahaa El Din, 1997). These publications adopt the rhetoric of globalization, modernization, and reform, with a focus on three areas. First, there is the discourse of *technology-based economic competition*: As noted by the TDC (1997),

> The whole world is undergoing an overwhelming technological revolution in information, electronics, computers, and communication. This revolution will widen the gap between the developed and underdeveloped countries. Those who master science and technology and manage information will survive, those who do not will perish, at least economically. Egypt must race against time so that it can jump on the wagon of the elite of the developed world before it is to late (p. 79).

The TDC goes on to explain that only through the infusion of modern technology in schools can this economic challenge be met.

Following on the heels of economic competition is the discourse of *educational transformation*. As Bahaa El Din (1997) writes,

> This emphasis [on technology] will have a transformative effect on education....The information explosion has changed education from a mode of memorization of a certain amount of knowledge to one in which students are expected to research and apply the knowledge they acquire to various life situations. Education will change from one that focuses on memorization to one that focuses on research, analysis, identification of relationships in the data, and potential application.

A sub-component of the discourse of educational transformation is that of autonomous learning. In MOE publications, multimedia laboratories, compact discs, the Internet, videoconference fiber optic networks, virtual reality, and electronic libraries will all provide learning resources so that students can engage in learner-centered experimentation, experiential learning, and critical thinking (Bahaa El Din, 1997; Technology Development Center, 1997).

Finally, the technology plans also emphasize equal opportunity for all. Distance education efforts, backed by the deployment of mobile technology caravans, are intended to bring educational resources to underserved students and thus bolster basic education and literacy (Technology Development Center, 1997).

The Government of Egypt (GOE) and MOE have assembled an impressive array of resources toward meeting these goals, including more than 600 full-time staff working for the TDC. The major technology projects involve computers and the Internet, satellite television, and video conferencing.

Computers and Internet: The TDC has placed multimedia rooms in all secondary and preparatory (i.e., middle) schools in Egypt and many primary schools. These rooms have 2-3 high-end computers, LCD devices for projecting from a computer to a screen, collections of educational software, and access to the Internet. These rooms are to be resources areas for teachers who can bring in their classes on a sign-up basis. Much of the school curriculum has been transferred to CD format for use in these multimedia rooms.

In addition to the multimedia rooms, secondary schools also have computer laboratories with 10-15 DOS or Windows computers. These courses are used for teaching an elective subject course called "computing" which is designed to cover basic operation and programming skills.

Satellite Television: Ten of the Nilesat television stations have been dedicated to educational program. Ministry of Education staff are creating educational television programs for seven of these stations based on the national curriculum. (The other three stations have been dedicated to the Ministry of Higher Education). Televisions, satellites, and digital receivers have been installed in the above-mentioned multimedia rooms in approximately ten thousand schools to facilitate access to the programming.

Videoconferencing Facilities: A national multipoint videoconference facility has been established, with videoconference training centers of 100-200 seats in each of Egypt's 27 governorates. The facilities are principally used for national teacher training programs and for national communications between Ministry of Education staff. The facilities allow participants in these programs to project from any site to all the other sites.

Practices of Educational Technology

The funds spent on information and communication technology represent a major investment for a developing country. What then are the results of this investment, and how do they match the MOE's lofty goals for technology in education? Unfortunately, results to date are unsatisfactory in all areas. Technology has been thrust on top of a mostly dysfunctional system, rather than used to help transform that system. The Technology Development Center itself is an add-on to the Ministry of Education that grabs up a huge portion of Ministry resources but appears to coordinate poorly with other sections of the Ministry, such as the departments of secondary or basic education or the department for inservice training. Serious problems have emerged in each of the three program areas:

Computers and Internet

The computers in the multimedia rooms, with 2-3 computers per school, seem to be spread too thin to make any difference. In any case, the rooms are often locked up, as local school authorities don't want to suffer the risk of having expensive equipment damaged. Classroom visitors representing donor agencies usually are given a special showcase presentation in a computer room. But, during those same visits, when I inspected the use logs, it was clear that many of the multimedia rooms in the schools I visited are rarely used outside of these formal visits. This phenomenon has frequently been reported often in the press. As one article (PCs and teachers omitted from new computer science curriculum, 2000) exclaimed,

> Primary School teacher Hasnaa el-Hefnawi is enraged by the decision to introduce the computer science curriculum...The ministry has repeatedly tooted its own horn about how many computers it has supplied to schools. "Doesn't the minister realize that these computers are kept in school warehouses like antiques or used merely for decoration" she mused (p. 2).

This sentiment was echoed by a teacher on an e-mail list of Egyptian educators, who complained about the technology gatekeepers at his own school, "And the good people know only how to unplug and cover it to protect the computer from dust so as not to be damaged."

During my visits to schools, when students did use these multimedia rooms, they usually sat and watched the teacher lecturing, as usual, but this time with the aid of a CD for presentation. The CDs themselves contain the exact same material as the textbooks, transferred to a new medium, with little attention given to principles of interactivity or participatory learning. Teachers who attempt to use the computers in more creative ways, even by making their own Microsoft Word or Microsoft PowerPoint files, have told me that they were warned that any activity other than using the Ministry-provided software is prohibited so as to protect against viruses.

Meanwhile, the laboratories of 10-15 computers are used for a course in basic computer literacy, which focuses for the most part on mastering DOS (or, in some cases, Windows) commands. Teachers of that class, as of other classes, told me that they are not allowed to depart from the prepared curriculum, nor are they prepared to do so based on knowledge, background, or training. The laboratories themselves, which could potentially offer a site for creative hands-on use by students in other subjects or after school, are generally forbidden to be used for anything other than the specified computer literacy courses, at least in the schools that I visited.

Finally, Internet access is routed by telephone via MOE Offices to ensure better control. This necessitates a double-connection process that rarely functions. In any case, in the schools that I visited, only the official in charge of the multimedia room was given the Internet account information, and neither classroom teachers nor students were allowed to access the Internet independently.

Satellite Television

The MOE rushed to transfer its entire curriculum to satellite television programming, similar to how it transferred the curriculum to CD format. In Egypt, the textbook *is* the curriculum, so this has too often meant simply converting an unappealing textbook into a similarly unappealing television program. Scriptwriters with more creative ideas have had their efforts rejected by the directors who are under pressure to develop an enormous amount of television material in a short amount of time. In any case, educational programming on satellite television appears to be rarely viewed, since relatively few people have bought a digital receiver at home and there is little reason to interrupt a class to bring students into a crowded television room to watch the same material that is found in their book.

Interactive Videoconferencing

The videoconference centers are used for teacher training, but the trainings that I have observed and heard about were more often based on lengthy talking head lectures from Cairo rather than real interaction. Scheduled videoconference trainings are frequently interrupted when the system breaks down or when top Ministry officials take over the system to communicate with subordinates around the country or to showcase the facilities to international visitors.

The ineffective use of videoconferencing parallels a broader problem with teacher training in new technology. Such training is generally reserved for the school computer specialists, and is generally limited to computer operations. The computer specialists have had no training in assisting teachers to make use of computers in teaching. Teachers themselves know little about either the pedagogy of instructional technology or even basic computer operations. As one university lecturer explained to me, "we have the hardware, we have the software, but we lack the *humanware*."

The problems with educational technology in Egypt are widely known and are reported frequently in the press (e.g., PCs and teachers omitted from new computer science curriculum, 2000). The ill-suited expenditures on technology—with the emphasis on hardware and software and inattention to promoting effective use of technology by skilled practitioners--serve to deepen public cynicism for the government and the Ministry of Education.

Educational Reform?

How then do these efforts stack up against Egypt's developmental goals of modernization, educational reform, and social inclusion? Though modernization and reform are the *raison d'être* of using technology in schools, the funds spent on technology have not served that purpose. Basic steps, such as using e-mail networks to facilitate coordination among teachers, have been ignored, in favor of high-profile but ill-suited expenditures. The Ministry rushes from one high-tech scheme to another, in recent years, rushing to transfer content to CD-ROMs, digital satellite television programs, and streaming video. In all cases, the content remains more or less the same, and the instruction is top-down, without engaging the type of interaction and inquiry among teachers and learners that the Ministry itself says is necessary for educational improvement (see, for example, Bahaa El Din 1997). The same top-down hierarchy permeates the TDC as other sections of the Ministry, giving classroom teachers—let alone students—little opportunity to exercise independent initiative.

In short, the curriculum, the exams, the teaching methods, and the need for expensive private tutoring have all remained the same. On a few occasions, ICT provides an alternate delivery mechanism, but the methods and content and approach to education have not substantially changed. ICT has not appeared to contribute in any meaningful way to reform and modernization of education.

There is also concern that ICT expenditures could be deepening social inequality. A major hindrance to Egypt's development is its unequal education system, and the resulting poor human capital development among the urban and rural poor, especially rural girls. The high rate of illiteracy in Egypt, especially among girls, is a major brake on development. Economists and development specialists believe that Egypt's educational expenditures are skewed to the well-off, and that Egypt spends an insufficient amount of its overall education pie on basic education and literacy promotion (Birdsall & Lesley, 1999; Institute of National Planning, 1998)

Not surprisingly, the investments in ICT have done little to overcome this bias and have likely worsened it. Egypt's investment in ICT, as in other countries, goes to those sectors best able to absorb it. When Egypt needs to be investing more in rural, primary education, ITC spending is skewed toward universities and secondary schools, which are located disproportionately in urban areas. With a mean years of schooling rate of 5.0 years (Fergany, 1998), much of the population never reaches the secondary schools that are absorbing the technological resources. A new effort to provide computers to university students at below-market prices is laudable on paper but will also put computing resources in the hands of those who can most afford them on their own. The expensive videoconferencing centers are based in governmental capital cities and draw money away from other types of school-based teacher training programs that could be spread more equitably around the country. In sum, the vast majority of spending on ICT is apparently going toward secondary and higher urban education, rather than toward improved primary rural education that could help combat illiteracy in Egypt.

In addition, an emphasis on ICT in education has tended to privilege the use of English over Arabic. Whereas textbooks available in Egyptian schools are all available in Arabic, much computing in Egypt takes place in English – due to English language computer science terms, English operating systems, English resources on the Internet, etc. (Warschauer et al., 2002). An increased emphasis on English—including the introduction of English in primary schools—has thus far borne little fruit (due to a lack of trained teachers, overcrowded classrooms, etc.), but has disadvantaged those students in rural primary schools who now have less time and opportunity to work toward gaining literacy in Arabic.

The Social Context of Educational Reform and Technology

It is not surprising that Egypt's educational technology effort has fallen short of its goals. Countries such as the United States, that have been spending a great deal more money on educational computing for a much longer period of time, are still far from getting it right. The learning curve for intelligent use of technology is a long and steep one, and there is no reason to expect Egypt to outperform other countries in this regard.

However, it is worth analyzing the Egyptian case in more detail to interpret the social context of educational technology difficulties. This may shed light on the broader issue mentioned at the beginning of this article as to whether the infusion of technology constitutes a lever for reform.

I believe the evidence of this study strongly supports the socialization view articulated by Cuban (1986; 1993a; 1993b) and others (e.g., Spindler, 1974) that gives priority emphasis to the broader social shaping role of schools. According to this view, deeply-held cultural beliefs about the nature of knowledge, how teaching should occur, and how children should learn steer policymakers and teachers toward certain forms of instruction, and that these forms of instruction are guided by the broader role of the schools to "inculcate into children the prevailing social norms, values, and behaviors that will prepare them for economic, social, and political participation in the larger culture" (Cuban 1993, p. 249). From this perspective, educational reform is not impossible, but tends to be available most often to the more privileged strata of society. Reforms affecting the masses are usually carried out in fringe ways, without disrupting the overall socialization function of the schools.

In this regard, it is useful to explore the broader social context that frames education in Egypt, and see how this framework constrains educational reform with computers. Three aspects will be examined: the *political*, the *cultural*, and the *economic*.

The Politics of School Reform

The political context of Egypt reflects a strong carry-over from the Nasser period, based on authoritarian rule by a military-backed leadership within a patriotic, nationalist framework (Hinnebusch, 1990). Egyptians enjoy neither freedom of speech, nor freedom of organization, nor freedom of organization and protest. Strikes and demonstrations are disallowed, those expressing contrary political or religious views are jailed, and formation of political parties and non-governmental organizations is restricted (The Economist Intelligence Unit, 1998-99). The current president, Hosni Mubarak, has been in power since 1981, and the country has been under Emergency Law during the entire time of his rule.

What then is the political role of schooling in Egypt, dating back to the Nasser regime? It is largely to forge a national identity based on mass access to (formally) equal schooling (see Jarrar & Massialas, 1992 for a history of Egypt's educational policies). Nasser brought huge numbers of children into the Egyptian school system, and construction of new schools continues to be a major priority of the current government (and deservedly so). However, dating back to the days of Nasser, any reform which allowed differences to emerge in schools, or which lessened the authoritarian hierarchy of the educational system, was highly suspect.

Today's political leaders, like Nasser, see schooling largely from the view of social control. Though the Islamist fundamentalist movement in Egypt is under greater control than it was in the 1990s, Islamist opposition remains a threat to the government, just as it has for the last 50 years (and, indeed, may grow due to public frustration with regional political events). In such a climate, a main function of schooling in Egypt, in the eyes of the regime, is to foster pro-government sentiment and to isolate the Egyptian fundamentalists. Toward this end, the *appearance* of modernization has proven very attractive. By constantly emphasizing how many computers it has put in schools and how advanced its videoconference system is, the government goes on the offensive to show that it represents the future and that it can compete with the wealthiest countries in the world. However, to actually *use* this equipment to reshape schooling would pose too much of a threat to a fundamentally conservative institution. So the system – beyond the wishes and desires of any particular individual – supports a rhetoric of reform without its substance. This is not due to the conspiracy of an elite, but rather due to the institutional reproduction of an educational and social system similar to that which occurs throughout the world.

The Culture of School Reform

For a variety of historical, political, social, economic, and religious reasons, a culture of vertical hierarchy permeates Egypt and the Arab world. Information is meant to be horded, decisions are made at the top, and rulers maintain power through a complex balance of power techniques. This hierarchical system, which de Atkine (2000) found to be evident in the Egyptian military, also pervades other social systems in the country (see discussion in Hudson, 2000). And indeed, an almost militaristic like atmosphere pervades the Ministry of Education, especially as it affects the use of technology. All three top leaders of the Technology Development Center are former military generals (none, by the way, with a background in education), and former corporals, lieutenants, and other officers are found below them. Computers are found on none of their desks, except as monitoring devices, i.e., to observe educational videoconferencing sessions organized within the Ministry. The TDC, like other governmental and MOE departments, is hierarchical to the extreme, with long chains of command, and those at any level but the top unable to make decisions. For example, on one occasion, I made a simple request of a teacher to see a copy of the CD that he uses in school. The request was bounced up one level after the other, with no one lower than the Vice-Minister of education willing to grant permission. (The Vice-Minister finally said yes.)

In such an atmosphere, it is not surprising that technology serves a purpose of hierarchy and transmission, rather than of horizontal networking (see discussion of this same issue in US education in Hodas, 1993). Though the MOE and TDC adopt the discourse of interactive education, the spending and support – whether on satellite television, or CDs, or top-down training via videoconferencing – has gone almost entirely to transmission technologies.

The Economics of School Reform

Egypt, like many developing countries, is highly stratified, yet that stratification is expressed in a special way. Due to the land reforms and other programs of the Nasser regime, *income inequality* and *land inequality* are relatively low. Education inequality, however, is quite high, even when compared to other developing countries. Data gathered by Birdsall and O'Connell (1999) illustrates this point (see table 1).

Table 1. Sources of Inequality Across Countries (Gini coefficients c. 1990)

Country	Income Inequality	Education Inequality	Land Inequality
Egypt	.320	.700	.480
Kenya	.544	.600	.746
Jordan	.407	.615	.686
Brazil	.596	.461	.852
Indonesia	.317	.494	.556

Korea	.336	.257	.351
Thailand	.515	.456	.366

Source: Birdsall and O'Connel, 1999

In other words, control of education—much more so than in other countries—is a principle means by which the well-to-do in Egypt defend their privileged social status. And the social status achieved through education means a great deal in Egypt. It allows people to join the highest paying and most prestigious professions, it is a de facto requirement for setting up your own business, and it is a prerequisite for marriage within the elite and thus enjoying the financial and social benefits showered on family members by the Egyptian upper class.

The defense of social status through elite education takes place through several means. First, a disproportionate share of funding goes to university education as opposed to K-12 education (Birdsall & Lesley, 1999; Fergany, 1998; Institute of National Planning, 1998). Secondly, socioeconomic privilege is reserved for those who complete the university; as mentioned earlier, those who graduate secondary school without a university education are actually worse off economically than primary school drop-outs unless they choose to work outside Egypt (Bartsch, 1995). Third, access to universities – and to the most elite departments or programs within the universities – is based on a set of decontextualized school-leaving exams that the wealthy have been preparing for all their lives, through their better private schools and their expensive private tutoring.

This class bias that permeates Egypt's educational system makes school reform extremely difficult to achieve. The Ministry of Education has attempted to disrupt the testing and private tutoring system for years, but has been continually rebuffed by a powerful elite who have invested huge sums in preparing their children to pass the exams and gain access to the elite (see Sarhaddi Nelson, 2001) and who thus have little interest in seeing such a system overturned.

This economic elite who exercise a powerful influence over Egyptian politics have little interest in technology-based school reform in governmental schools. Their children are already becoming computer-proficient at home, and for them the schools serve as little more than a sorting system to maintain their class privilege.

Meanwhile the poor have little vested interest in demanding computers in the schools. The struggle of the poor is for decent basic education that will allow their children to read and write and compete fairly in society. With class sizes of upwards of 60 in the poor neighborhoods of Cairo, and many urban and rural schools lacking basic amenities, the poor have other priorities than computers and the Internet in schools, which are widely viewed as a boondoggle. There is thus no constituency that is fighting hard to reform schools through infusions of technology.

In summary then, there are powerful political, cultural, and economic factors motivating the current structure of education in Egypt. Large-scale spending on information technology has had little if any impact on changing these factors, and it is unrealistic to expect that it will. The emphasis on the façade of reform without any substantial changes is evident throughout the educational system. The Ministry began teaching English in elementary school in order to emphasize modernization and ties to the West, but this instruction has almost no value because the majority of those designated to teach English in primary school know little of the language itself. (Indeed, a committee of Egyptian applied linguistics, several with expertise in the field of English language teaching, recommended against the change to earlier English language education.) The Ministry, at huge expense and often with the support of donor funding, also sends thousands of teachers per year to the United States and Britain to expose them to Western environments and approaches, but makes it difficult for these same teachers to implement any substantial changes when they return (Warschauer, 2003b). In other words, whether in English teaching, teacher training, or use of technology, a higher priority is put on creating the illusion of modernization rather than on actually modernizing practices.

Finally, though beyond the scope of this particular work, the role of donor agencies must also be briefly mentioned. For example, the United States Agency for International Development has long made Egypt one of its largest recipients, principally for global political reasons (Weinbaum, 1986). In too many cases, the US and other donors have poured money into expensive infrastructure projects in Egypt, including those related to technology in education, without paying sufficient attention to how technology might actually be best used in local contexts (see discussion in Warschauer, 2003a, 2003b).

Conclusion

While the large gap between rhetoric and reality of technology-based educational reform in Egypt stems in part from poor planning, it is also the logical outcome of powerful socioeconomic factors that shape educational policy and practice in Egypt. Though Egyptian officials voice the discourses of reform popularized in the West, it is unlikely that they will be widely practiced in Egypt (and, indeed, it is questionable how often they are actually practiced in Western countries.) Egypt would do better to draw on its own social norms in designing educational reform policies. Holliday (1992; 1994) has demonstrated that the best classroom instruction in Egypt is teacher-centered, reflecting the social and cultural realities of Egypt rather than the learner-centered environment favored in US graduate schools of education; most Egyptians believe that their educators should organize "teaching spectacles" featuring top-down instruction rather than "learning festivals" based on collaborative project-work (Holliday, 1994, p. 36). If and when educational technology begins to make positive headway in Egypt, it will be in a way that is likely very different than a Western-based discourse might suggest. Indeed, the US has its own share of educational problems, and many countries with traditions of teacher-centered education, such as Korea, Singapore, and Taiwan, have shown great success in raising the educational level of their students.

As for the more general conclusion, this study provides support for the work of Cuban and others who emphasize the limited impact of machines in reforming education. In Egypt, as in the US, technology can play a role in remaking education only if an when broader social, political, cultural, and economic factors are aligned to make school reform likely. This does not suggest a fatalistic approach that denies human agency, but it does imply that technology is something other than a neutral tool that can be deployed toward any ends. It is better to think of information and communication technologies as "socio-technical networks" (Kling, 2000, p. 3) that involve complex social relationships and contexts.

In other words, those who seek to reshape schools, whether in Egypt, the US, or elsewhere, need to think about not only the technology of the classroom, but also the technology of informational capitalism. Globalization, post-Fordism industrial relations, and the advent of new communications media are changing the context of education in the US, other Western countries, and, increasingly, in the developing world. These broader economic shifts may well introduce a greater demand for a more educated (or differently-educated) workforce in Egypt, and there are already signs that the business community in Egypt is starting to throw its weight toward educational reform for just this reason. Modernization of the educational system in Egypt, as elsewhere, will come about because influential social forces push for it, not because x number of computers have been put in y number of schools. Changes in the political economy can result in a context that better supports reform, but even then reform will not happen on its own. Working for educational reform requires not *machines* but rather *mobilization*—that is, the engagement of social actors to press for change, taking into account the relevant political, economic, and cultural contexts that help shape classroom learning and teaching.

References

Agre, P. (1997). Criando uma cultura da Internet. *Revista USP (University of Sao Paulo), 35*, 112-117.

Badawi, A. (2000, July). Telecommunications in Egypt: Presentation at the US Embassy, Cairo.

Bahaa El Din, H. K. (1997). *Education and the future*. Kayoub, Egypt: Al Ahram Commercial Press.

Bartsch, U. (1995). *Rates of return to investment in education and migration in Egypt* (Research Notes 9). Cairo: Almishkat Cenre for Researcha and Training.

Becker, H. (2000). Who's wired and who's not? *The future of children, 10*(2), 44-75 (available at http://www.futureofchildren.org/cct/cct_03.pdf).

Becker, H. J. (1982). *Microcomputers in the classroom: Dreams and realities (Report No. 319)*. Baltimore, MD: John Hopkins University, Center for Social Organization of Schools.

Becker, H. J. (2000). Findings from the Teaching, Learning, and Computing survye: Is Larry Cuban Right. *Educational Policy Analysis Archives, 8*(51).

Birdsall, N., & Lesley, O. C. (1999). *Globalization, income distribution and education: Putting education to work in Egypt*. Cairo, Egypt: Egyptian Center for Economic Studies.

Calderoni, J. (1998). *Telesecundaria: Using TV to bring education to rural Mexico* (Education and Technology Technical Notes Series 3(2)). Washington, DC: World Bank.

Castells, M. (1996). *The rise of the network society*. Malden, MA: Blackwell.

Castells, M. (1997). *The power of identity*. Malden, MA: Blackwell.

Castells, M. (1998). *End of Millennium*. Malden, MA: Blackwell.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College Press.

Cuban, L. (1993a). Computer meets classroom: Classroom wins. *Teachers College Record, 95*(2), 185-210.

Cuban, L. (1993b). *How teachers taught: Constancy and change in American classrooms 1890-1980* (2nd ed.). New York: Longman.

de Atkine, N. (2000). Why Arabs lose wars. *Middle East Review of International Affairs, 4*(1), 16-27.

Arab Communication Consult (2002). *Egypt Investment Report 2002*. [Online report.] Downloaded June 6, 2003 from http://www.arabcomconsult.com/egyptr/food.htm.

El-Nawawy, M. A. (2000). Profiling Internet users in Egypt: Understanding the primary deterrent against their growth in number, *INET 2000 Proceedings*. Downloaded March 2, 2001 from http://www.isoc.org/inet2000/cdproceedings/8d/8d_3.htm.

Ellul, J. (1980). *The technological system*. New York: Continuum.

Feenberg, A. (1991). *Critical theory of technology*. New York: Oxford University Press.

Feldman, A., Konold, C., & Coulter, B. (2000). *Network science, a decade later: The Internet and classroom learning*. Malwah, New Jersey: Lawrence Erlbaum Associates.

Fergany, N. (1998). *Human capital and economic performance in Egypt*. Cairo: Mimeo.

Fergany, N., Farmaz, I., & Wissa, C. (1996). *Enrollment in primary education and cognitive achievement in Egypt: Change and determinants*. Cairo: ALMISHKAT Cenre for Research and Training.

Galal, A. (1995). *Which institutions constrain economic growth in Egypt?* (Working Paper Series 1). Cairo: Egyptian Center for Economic Studies.

Harnad, S. (1991). Post-Gutenberg galaxy: The fourth revolution in the means of production and knowledge. *Public Access Computer Systems Review, 2*(1), 39-53.

Hinnebusch, R. A. (1990). The formation of the contemporary Egyptian state from Nasser and Sadat to Mubarak. In I. M. Oweiss (Ed.), *The political economy of contemporary Egypt* (pp. 188-209). Washington, DC: Center for Contemporary Arab Studies, Georgetown University.

Hodas, S. (1993). *Technology refusal and the organizational culture of schools*.

Holliday, A. (1992). Tissue rejection and informal orders in ELT projects: Collecting the right information. *Applied Linguistics, 13*(4), 404-424.

Holliday, A. (1994). *Appropriate methodology in social context*. Cambridge: Cambridge University Press.

Hudson, M. (2000). A "Pan-Arab Virtual Think Tank": Encircling the Arab information environment. *Middle East Journal, 54*(3).

Information technology in Egypt. (1998). Cairo: American Chamber of Commerce in Egypt.

Institute of National Planning. (1998). *Egypt: Human Development Report 1997-1998*. Cairo: Institute of National Planning.

Jarrar, S. A., & Massialas, B. G. (1992). Arab Republic of Egypt. In J. Cookson, Peter W. & A. R. Sadovnik & S. F. Semel (Eds.), *International handbook of educational reform*. New York: Greenwood Press.

Kling, R. (2000). Learning about information technologies and social change: The contribution of social informatics. *The Information Society, 16*(3), 1-36.

Knapp, L. R., & Glenn, A. D. (1996). *Restructuring schools with technology*. Needham Heights, MA: Allyn and Bacon.

Korayem, K. (1997). *Egypt's Economic Reform and Structural Program (ERSAP)*. Cairo: Egyptian Center for Economic Studies.

Means, B. (1998). Models and prospects for bringing technology-suported education reform to scale. In April (Ed.), *American Educational Research Association Annual Meeting*. San Diego.

Mehlinger, H. D. (1996). School reform in the information age. *Phi Delta Kappan, 77*, 400-407.

Ministry of Education. (1993). *General report on National Survey of Teaching Practices, Sudent Achievement, and School Effectiveness*. Cairo: Eduation Planning and Information Division, Research and Analysis Directorate, Ministry of Education.

Ministry of Education. (1999). *Mubarak and education: The national project for developing education*. Cairo: Egypt.

Mintz, S. (1999). *The Internet as a tool for Egypt's economic growth*. Burke, VA: Report for the United States Agency for International Development.

New and old: A survey of Egypt. (1999, March 20-26). *Economist, 350*, Special survey 1-18.

NUA. (2003). *How many online?: Africa*. Downloaded January 15, 2003 from the World Wide Web: [http: //www.nua.ie/surveys/how_many_online/africa.html]

Osin, L. (1998). *Computers in education in developing countries: Why and how?* (Education and Technology Series Volume3, Number 1). Washington, DC: World Bank Education and Technology Team.

PCs and teachers omitted from new computer science curriculum. (2000, September 22).*Egyptian Gazette*, pp. 2.

Potashnik, M. (1996). *Chile's learning network* (Education and Technology Series Volume 1, Number 2). Washington, DC: World Bank Education and Technology Team.

Sachs, J. (1996). *Achieving rapid growth: The road ahead for Egypt*. Cairo: Egyptian Center for Economic Studies.

Sakr, N. (1999). *Satellite television and development in the Middle East* (Middle East Report 210): Middle East Research and Information Project.

Sandholtz, J. H., Ringstaff, C., & Dwyer, D. C. (1997). *Teaching with technology: Creating student-centered classrooms*. New York: Teachers College Press.

Sarhaddi Nelson, S. (2001, Jnuary 21). Egyptians' obsession with grades fails to nurture creative thinkers. *Los Angeles Times*, pp. A8.

Shallis, M. (1984). *The silicon idol: The Microsoft revolution and its social implications*. Oxford: Oxford University Press.

Singh, J. P. (1999). *Leapfrogging development?: The political economy of telecommunications restructuring*. State University of New York Press: Albany.

Spindler, G. D. (1974). The transmission of culture. In G. D. Spindlier (Ed.), *Educational and cultural process: Toward an anthropology of education*. New York: Holt, Rinehart and Winston.

Starr, P. (1996). *Computing our way to educational reform.*

Street, B. (1984). *Literacy in theory and practice*. Cambridge: Cambridge University Press.

Street, B. (1993). Introduction: The new literacy studies. In B. V. Street (Ed.), *Cross-cultural approaches to literacy* (pp. 1-21). Cambridge: Cambridge University Press.

Tawila, S., Lloyd, C. B., Bensch, B. S., & Wassef, H. (2000). *The school environment in Egypt: A situational analysis of public preparatory schools*. Cairo: Population Council.

Technology Development Center. (1997). *Technology: A tool for developing education in the 21st century*. Cairo: Ministry of Education.

The Economist Intelligence Unit. (1998-99). *Country profile: Egypt*. London, England: Author.

United Nations Development Programme. (2000). *Human development report 2000*. New York: Oxford University Press.

Warschauer, M. (1998). Researching technology in TESOL: Determinist, instrumental, and critical approaches. *TESOL Quarterly, 32* (4), 757-761.

Warschauer, M. (1999). *Electronic literacies: Language, culture, and power in online education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Warschauer, M. (2000). Technology and school reform: A view from both sides of the track. *Education Policy Analysis Archives, 8*(4). Retrieved October 1, 2003 from http://epaa.asu.edu/epaa /v8n4.html.

Warschauer, M. (2003a). *Technology and social inclusion: Rethinking the digital divide*. Cambridge, MA: MIT Press.

Warschauer, M. (2003b, April). *The rhetoric and reality of aid: A critical look at shared responsibility*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago. Downloaded June 6, 2003 from http://www.gse.uci.edu/markw/rhetoric.pdf.

Warschauer, M., El Said, G. R., & Zohry, A. (2002). Language choice online: Globalization and identity in Egypt. *Journal of Computer-Mediated Communication, 7*(4). Downloaded January 17, 2003 from http ://www.ascusc.org/jcmc/vol7/issue4/warschauer.html.

Weinbaum, M. (1986). *Egypt and the politics of U.S. economic aid*. Boulder, CO: Westview.

World Bank. (2001). *World development report 200/2001: Attacking poverty*. Oxford: Oxford University Press.

About the Author

Mark Warschauer is Vice Chair and Associate Professor of the Department of Education at the University of California, Irvine and the editor of *Language Learning & Technology* journal. His research focuses on the uses of information and communication technologies with culturally and linguistically diverse learners. His most recent book is *Technology and Social Inclusion: Rethinking the Digital Divide* (MIT Press, 2003). He can be reached through his Website at http://www.gse.uci.edu/markw.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

Sherman Dorn
University of South Florida

Gustavo E. Fischman
Arizona State Univeristy

Thomas F. Green
Syracuse University

Craig B. Howley
Appalachia Educational Laboratory

Patricia Fey Jarvis
Seattle, Washington

Benjamin Levin
University of Manitoba

Les McLean
University of Toronto

Michele Moses
Arizona State University

Anthony G. Rud Jr.
Purdue University

Michael Scriven
University of Auckland

Robert E. Stake
University of Illinois—UC

Terrence G. Wiley
Arizona State University

David C. Berliner
Arizona State University

Linda Darling-Hammond
Stanford University

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
Birmingham, Alabama

Aimee Howley
Ohio University

William Hunter
University of Ontario Institute of Technology

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

Heinrich Mintrop
University of California, Los Angeles

Gary Orfield
Harvard University

Jay Paredes Scribner
University of Missouri

Lorrie A. Shepard
University of Colorado, Boulder

Kevin Welner
University of Colorado, Boulder

John Willinsky
University of British Columbia

EPAA Spanish & Portuguese Language Editorial Board

Associate Editors for Spanish & Portuguese

Gustavo E. Fischman
Arizona State University
fischman@asu.edu

Pablo Gentili
Laboratório de Políticas Públicas
Universidade do Estado do Rio de Janeiro

pablo@lpp-uerj.net

Founding Associate Editor for Spanish Language (1998-2003)
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

home abstracts complete editors submit comment notices search

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES**. **EPAA** is a project of the Education Policy Studies Laboratory.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

PDF file of this article available:

This article has been retrieved 1336 times since October 20, 2003

Volume 11 Number 39 **October 20, 2003** **ISSN 1068-2341**

Baselines for Assessment of Choice Programs

Paul T. Hill
University of Washington

Kacey Guin
University of Washington

Citation: Hill, P. T. & Guin, K. (2003, October 20). Baselines for assessment of choice programs. *Education Policy Analysis Archives, 11*(39). Retrieved [Date] from http://epaa.asu.edu/epaa/v11n39/.

Abstract

Critics of choice argue that it will allow alert and aggressive parents to get the best of everything for their children, leaving poor and minority children concentrated in the worst schools. (Note 1) But choice is not the only mechanism whereby this occurs. Alert and aggressive parents work the bureaucracy to get the best for their children. Thus, choice programs should be compared against the real performance of the current public education system, not its idealized aspirations.

The purpose of this article is to establish an appropriate baseline against which choice programs can be assessed. How far does the current system of bureaucratic allocation diverge from its aspirations for equal opportunity for all? Under the current system how much are students sorted by race and class, and how unevenly allocated are the best and worst educational experiences? The answers to these questions are important for two reasons:

> First, they establish defensible baselines against which choice programs can be compared. If the current ways of allocating educational opportunities leads to inequality by race, class, or income, then choice programs should not be assessed against the ideal. Instead, their results should be compared to the actual performance of the existing system.
>
> Second, they establish criteria for the design of choice programs. Regardless of whether choice programs are on average better or worse than other ways of allocating educational programs, there are still ethical reasons for trying to design choice schemes to equalize access to the best schools and teachers.

Bureaucratic modes of decision-making do not eliminate self-seeking – they only make it covert. When the supply of desirable schools, programs, or teachers is limited the most aggressive get the best and by implication deprive others. In bureaucracies, the advantage goes to people who have contacts, understand how the game is played, can talk the language of key administrators, can write letters and threaten appeals, and have the time and determination to persist. These attributes have a strong class bias. As a result, bureaucratic decision-making can create segregation of students and uneven distribution of benefits. These, of course, are the very outcomes that people fear choice will produce.

Choice is another mechanism by which people seek the best for themselves and their children. The most knowledgeable are first to identify the best opportunities, and the most aggressive are the ones most likely to sign up early, know how to get the most advantageous place in a lottery, and be able to impress people (e.g. admissions officers) who can pick from among many applicants.

Self-seeking would not matter if all schools, teachers, or courses were equally good. But that is not the case. To the contrary, some schools are much better than others, even when quality is measured fairly on the basis of what they add to their students' knowledge. (Note 2) There is also reason to believe that some teachers are much better than others(Note 3) and also that some courses of study are much more likely to prepare students for jobs and higher education than others. (Note 4) Because some students thrive in schools that would not be good for other students, there is more than one way to rank quality. But however quality is defined, the "best" schools and teachers are usually in short supply. That is why the most respected private schools have long waiting lists and why parents camp out in parking lots to register their children in public magnet schools.

Some public school districts try to provide a quality school for every student, but they are thwarted by scarcity. There are only so many experienced teachers, only so many principals who can create a positive school climate, and only so many people who both understand science and mathematics and want to teach those subjects. Schools are like any other enterprise that depends on people. Only so many children can take chemistry from the fabled teacher whose students regularly end up in medical school. Someone will get the burned out old teacher in his last year, or the brand new teacher whose command of subject matter and classroom management skills is shaky. Some schools or districts might maximize the average quality of their staffs, encourage the burned-out to retire earlier, or do a better job of mentoring inexperienced teachers. But there will always be differences in quality, both real and perceived.

Scarcity begets competition. Though some parents will knowingly accept less than the best for their children, many will not. Among those who try to get the best (or to spare their children contact with the worst), some will fare better than others. Those who do not try to compete will probably do worse than even the least successful competitors. (Note 5)

How people compete for schools and teachers depends on the way opportunities are allocated. When parents are free to apply to any public school, the most competitive study the options, apply early, and try to make sure they apply to some desirable schools where the probability of admission is high. When parents are assigned to schools, the most competitive learn who are the best and worst teachers and programs and campaign to get these for their children. (Note 6) , (Note 7)

The rules of competition inevitably allocate advantages and disadvantages. When the rules allow exceptions to mandatory school assignment, families eager to get the best for their children learn how decisions are made and frame their transfer appeals in then appropriate terms. They also figure out who makes the final decision on transfer requests, and seek ways to get consideration. Thus, choice is only one way of allocating educational opportunities. Self-seeking and competition are universal. Only the means differ.

The advantage of choice is that advantage seeking is transparent, its effects can be readily observed, and it can be designed out (e.g. via admissions lotteries). Self-seeking in bureaucracies is covert and is therefore harder to observe and remedy.

Whether choice or bureaucratic decision-making lead to a "fairer" allocation of opportunities is an empirical question. Under both systems, the advantaged are likely to get a disproportionate share of the best and the disadvantaged are likely to get the worst. Thus the question for public debate is not whether choice leads to inequalities but whether it leads to any greater inequalities than does non-choice.

Perhaps a better way to formulate this question is whether overt choice leads to the same or lesser inequities than does covert choice. As David Menefee-Libey of Pomona College has suggested, someone always exercises choices, even in bureaucratic systems. What matters is whether everyone or just some people have choices, and whether choices are made openly or in secret. Overt school choice occurs when everyone can choose and everyone who picks a particular school has an equal chance of getting in. Covert choice occurs when there are no structured mechanisms for expressing choices and allocating opportunities, so that families who want particular options are forced to campaign for them. Because families must go out of their way to express choices, and must work the bureaucracy to get what they want, covert choice strongly favors the sophisticated and well placed.

Critics of overt education choice proposals assert that they make matters worse for the disadvantaged and promote development of privileged enclaves for the advantaged. The implication of these statements is that choice makes things worse than they are now. But the evidence provided is often quite different. It shows that overt choice leads to some unequal outcomes, but it does not show that choice leads to more unequal outcomes than the covert choice system that now prevails.

Establishing a Baseline

Critics claim that choice will worsen segregation and other forms of inequity. This article asks, compared to what? The proper baseline against which to assess the effects of choice is the current system's performance, not some idealized situation in which no differences exist. As Stephen Gorard and his colleagues observe about universal choice in Britain, "The stratifying effect of market forces in schools depends, to large extent, on the *status ante*. What we have shown is not that choice is SES-free but that it is certainly no worse, and probably a great deal better, than simply assigning children to their nearest school to be educated with similar children living in similar housing conditions." (Note 8)

Using the current system's performance as a baseline for comparison does not imply satisfaction with things as they are. Programs that rely on choice should be designed to produce less segregation and more equitable distributions of resources and opportunity than now exist. This article, however, focuses narrowly on whether defenders of the current system are justified in opposing choice on grounds that it inevitably worsens segregation and inequitable distribution of resources. Our narrow question is this: if public funds were used to create many options for families, and families were free to choose among those options, would segregation and inequity be worse than it is now?

We provide a baseline of evidence by which the consequences of choice can be compared with the results of the current public school system. Some critics of choice would like to compare it against an idealized form of the current system: Gary Orfield, among others, asserts that the current system can be perfected to eliminate any form of segregation, even those based on residential choices. He argues for "deny[ing privileged families] the possibility of finding nearby all-white schools," (Note 9) via creation of metropolitan-wide school districts, and massive busing to ensure racial mixing in all schools regardless of residential segregation. It is beyond the scope of this paper to assess the political, legal, and financial costs of such a scheme, or its implications for the health and education of children.

In establishing a baseline we will focus on the sorting effects of several bureaucratic processes endemic to conventional public school systems. (Note 10) These include student assignment and resource allocation processes that lead to disadvantaged children to experience:

- Racially isolated schools;
- Less money per pupil and less capable teachers;
- Restricted access to instructional programs that enhance life opportunities;
- Enhanced access to instructional programs that limit life opportunities.

The article has two main sections. The first main section immediately below analyzes the ways that each of the harms listed above can occur in conventional public school systems, and summaries of the available evidence about how often and how severely these harms actually occur.

The second main section identifies they ways in which these same harms can occur under choice programs, and summarizes available evidence about the performance of choice programs. This section is inevitably weakly-evidenced and tentative, since existing choice programs are small and often designed to serve the poor and ensure integration. Universal choice programs (in which every family chooses and every school is a school of choice) might work differently than the exemplars available for study today.

The Harms of the Existing System

Racially Isolated Schools

Eliminating segregation by race has been a dominant concern of public school systems since the *Brown* decision in 1954. Every large school system has had a desegregation plan, whether court-ordered or voluntary, and the U.S. Department of Education has monitored racial isolation in every school district large and small. No school district has an overt segregation policy, and most have made significant efforts to create racially mixed student bodies. However, as we will discuss immediately below, most districts remain segregated to some degree, and segregation has recently increased. (Note 11)

How does this happen? In part it happens because of processes that school systems do not control: housing economics, demographic change, and geography. Low-income families, including the majority of Hispanic and African American households, cluster in neighborhoods with low-cost housing. Wealthier families, most of which are white, avoid living in these neighborhoods. Lower-income minority families also have more children than higher-income white families. This leads to concentrations of minority children in certain neighborhoods. (Note 12) In many cities (e.g. Seattle) transportation between white and minority neighborhoods is complicated by bridges and choked freeways, making it very difficult to move children from one neighborhood to another.

Public school systems can exacerbate these problems by maintaining attendance boundaries that divide neighboring minority and white areas. They can respond to growing minority enrollments by enlarging schools deep in minority areas rather than by developing new schools in areas accessible to people of all races. They can also create admissions processes for attractive magnet schools that give the advantage to aggressive, articulate, and well-connected middle class parents. Finally, they can limit the supply of schools that students from all neighborhoods want to attend, e.g. by maintaining a fixed set of schools rather than expanding or duplicating magnet schools that have long waiting lists. Taken together these actions can lead to significant segregation by race and ethnicity.

The Baseline Level of Racially Isolated Schooling

Reports from the Harvard Project on Civil Rights provide data on segregation nationwide. One simple measure is the proportion of white students in schools attended by students of different races. In 1999, the school attended by a typical white student was 81.2% white, 8.6% African-American, 6.6% Latino, 2.8% Asian, and 0,8% American Indian. In contrast, the school, attended by a typical African American student was 32.6% white and 54.5% black. Latinos were even more segregated: the typical Latino student attended a school that was only 29.9% white. (Note 13)

Though school segregation has decreased markedly since 1960, separation of white and minority students has increased since 1988. In the South, as Orfield reports, the proportion of black students enrolled in majority white students declined from 43.5% in 1988 to 32.7% in 1998. (Note 14)

Much of the recent growth in segregation has been caused by a decline in the numbers of white students in the schools (from 34.7 million in 1988 to 28.9 million in 1998) and growth in the numbers of minority students (from 8.3 million to 14.8 million in the same period). (Note 15) Changes have been most dramatic in the west, where whites went from 63.3% of public school enrollment to 51.9% in the 11-year period between 1987-1998. Many big cities have also become minority enclaves. In 1998, white students made up less than 20% of the public school population in 18 of the 25 largest cities. Schools in Chicago Detroit, Dallas, New Orleans, DC and Atlanta are no more than 10% white. (Note 16) Thus, in some localities there is no way to avoid having some overwhelmingly minority schools.

Segregation is pronounced even in states with few minority students. For example, in 1998, the typical black student in a state in which only 1 in 16 students was black is likely to attend a school in which more than 1 in 2 students was black. (Note 17) Nationwide, black students, who made up only 18% of the school population in 1998, had a 37% chance of going to schools where blacks made up more than 90% of the student body. (Note 18)

Though data on individual school districts can be hard to find, racial isolation is common. In Louisville, for example, black students make up 27.4 % of the high school population, but 6 of 20 high schools have student bodies less than 20% black and 6 have student bodies more than 40% black. (Note 19) In Charlotte-Mecklenburg, like Louisville a city in which white students are in the majority (54%), 27% of white students and18% of blacks were in racially isolated schools. Under Charlotte's court-ordered definition, a white student is in a racially isolated school if its population is more than 69% white. The corresponding number for black students is 56% black). (Note 20) In Charlotte, more than 35% of public schools are racially isolated under the local definition. In a much more racially unbalanced city, the District of Columbia, whites are less than 4.3% of the school population. However, the average white student attends a school where the combined black and Latino population is less than 50%. (Note 21)

Individual school districts will vary, but these underlying facts reflect a common pattern. They set a baseline against which the segregation effects of choice can be measured. Choice programs might lead to worse segregation than we now have – to a situation where, for example, where blacks nationwide have a greater than 50% chance of attending schools that are more than 90% black, or where the average white student goes to a school in which even less than 20% of students are black. However, as these data show, the existing system does not live up to its rhetorical commitment to complete racial mixing. Choice programs should surely be compared against the system's real performance, not its aspirations.

Dollar and Human Resource Inequities

Public school districts receive funds from many sources – local property taxes, their state's basic school funding formula, various state programs that provide money for defined purposes and various federal funding sources – and the districts use these funds in similarly complex ways. Laypersons might expect that money is allocated to schools on a per-pupil basis, but that is not the case. Districts buy things like teachers, books, equipment, expert advice, buses, school construction, and maintenance, and those things are allocated to schools via political and bureaucratic processes. The result can be that some schools get the benefit of much higher spending, and receive much more valuable resources, than others.

The most valuable resources allocated in this way are teachers. In virtually all school districts, teachers allocate themselves to schools, and the most senior and highest-paid teachers get first choice. The majority of senior teachers choose schools in the "nicer" neighborhoods. The result is that the teachers who work in schools with the most advantaged students are, on average, much higher-paid than teachers who work in the poorer ends of town. The poorer students are not compensated for this difference in average teacher salaries. Instead, the district's public accounts average out the salaries of all teachers so it does not look like the schools with many expensive senior teachers have any more money than the schools with many cheap new teachers. On a real-dollar basis, per pupil expenditures are much higher in the schools chosen by senior teachers.

Though staff salaries constitute as much as 80% of school-level expenditure, districts allocate other resources to schools. Poor schools get disproportionate shares of the 10% of funds that come from federal and state programs intended for low-income students. This does only a little to compensate for the expenditure differences associated with teacher allocation.

Funds for the education of children with disabilities are allocated on the basis of diagnoses of children's needs. Parent initiative is a major factor in children's diagnoses: more sophisticated parents demand and get expensive individualized placements for their children with disabilities, while less sophisticated parents are less likely to take the initiative. Low-income and minority children identified with disabilities are therefore much more likely to be assigned to self-contained special education classrooms for mental retardation or emotionally disturbance than to be "mainstreamed" in general education classrooms and receive related services. (Note 22)

Districts also control other resources, from computers and science lab equipment to maintenance work, and these are allocated on a "squeaky wheel" basis. Schools with respected principals and teachers, and with active and well-connected parents, can capture disproportionate shares of these resources.

Though district accounting makes it extremely difficult to compute real-dollar per-pupil expenditures, within-district resource allocation consistently favors the more aggressive and influential families and neighborhoods.

The Baseline Level of Resource Inequity.

The existing system allocates the two most important resources in education – dollars and quality teachers – by bureaucratic means. The result is dramatic inequity within school districts. (Note 23)

Analyzing school funding in Seattle, Cincinnati, and Houston, Marguerite Roza found that some elementary schools in poverty neighborhoods received real-dollar resources worth as much as $300,000 less than was claimed by the district's budget, and that similarly-sized schools in high-income neighborhoods got correspondingly more money that the district budget claimed. This was caused by a combination of placement privileges for senior teachers – which allow senior teachers to cluster in schools in higher-income neighborhoods – and average teacher costing, which charges schools the same amount for every teacher whatever that teacher's actual salary. Under such a scheme schools in nice neighborhoods get a more expensive teaching force than they could afford if they paid real prices for teachers, and schools in poorer neighborhoods get a much cheaper teaching force. (Note 24)

When Houston school officials computed real-dollar spending in their high schools they were shocked to learn that one school in a predominantly white section of town had one million dollars more to spend each year than a school of the same size in a minority area. The difference, they learned, was entirely due to differences in teacher pay. Teachers in the higher spending white school were older and more experienced. (Note 25)

It is important to note that Seattle, Cincinnati, and Houston are not isolated incidents when it comes to inequalities in school funding. State-by-state data from The Education Trust indicate that schools with a high percentage of low-income students receive anywhere from $32 to $2700 less per student than schools with a low percentage of low-income students. (Note 26) A disparity in funding was found in 42 out of the 49 states studied. (Note 27)

Access to qualified teachers also produces inequalities between racial and socio-economic groups. In California, the number of economically disadvantaged students in a school is positively correlated with the number of teachers having the least amount of teaching experience and holding a bachelor's degree or less. (Note 28) This correlation is particularly strong in the elementary grades. In secondary education, national data indicate that 25% of classes in high poverty schools are taught by teachers lacking a major or minor in the field they teach, compared to 15% of classes in low poverty schools. (Note 29) This disparity is even greater for math, where only 25% of the teachers in high poverty schools were majors in math, compare to 40% of higher income schools. (Note 30)

Inequalities also exist based on schools' racial composition. In schools where the student population is over 90% white, 69% of teachers have BAs or higher in math versus 42% in schools where 90% or more of students are minority group members. (Note 31) National data show similar disparities, with 22% of teachers in high minority secondary schools lacking a major or minor in the field they teach, compared to 16% of teachers in low minority schools. (Note 32)

When examining the differences in human resources among schools, it is important to address the negative results of ineffective teachers. These results can are found at both the elementary and secondary levels. In Dallas, 5th grade students who had three consecutive ineffective teachers showed gains of only 29% in math scores, compared with an 83% gain for students with three years of effective teachers. In Boston, high school students had average gains of –0.6 in math and 0.3 in reading after one year with ineffective teachers, compared to students with effective teachers, who had average gains of 14.6 and 5.6 respectively.

Allocation of Opportunity-Limiting Programs

The fact that students come to school – any school – with different amounts of prior knowledge and different abilities presents problems for teachers, schools, and districts. (Note 33) Teachers find it difficult to prepare lessons and oversee learning for students with very diverse prior experiences and ability. Parents of the more advanced students worry that teaching will be tailored to the needs of others, and that their children will consequently learn less than they might. Parents of the less advanced students are also forthright in demanding that their children get extra help and attention.

The response by public schools and school districts is to differentiate instruction and create homogeneous classroom groups. The federal and state governments also provide special funding for instruction for defined groups, especially low-achieving students, children in poverty, and the handicapped.

Some differentiation of instruction is inevitable and some might be desirable. But there are ways in which it can harm minority and disadvantaged students. Removing students from regular classrooms to get special drills and tutoring can mean that they never master the material that others learn while they are away. (Note 34) (Note 35) Reducing contact with advanced students can eliminate a potential learning opportunity. Creating programs that focus on low-level skills can discourage children who are excited about ideas and could be motivated by highly challenging instruction. Creating a low-status program might discourage both students and teachers and set off a downward spiral of expectations and performance.

There is a raging debate about the educational value and ethical acceptability of the combination of ability grouping and program differentiation. (Note 36) But there is little dispute about the fact that some students are assigned to such programs on the basis of color and family background, and that there can be significant overlaps in the ability of students assigned to less and more challenging programs. Nor is there any doubt that program assignment affects students' likelihood of completing high school. The current system, by the way it designs special instructional programs and assigns students to them, puts some students at a grave disadvantage.

The Baseline Allocation Of Opportunity-Limiting Programs

UCLA education researcher Jeannie Oakes is the most important source of data on the assignment of students to opportunity-limiting courses, called tracking. In her 1985 book *Keeping Track* (Note 37) she shows that schools with different instructional programs for students considered faster and slower consistently assign minority and low-income students to the slower tracks. Though track placement is meant to correlate with student performance on achievement tests and grades in previous classes, Oakes reports significant overlap in ability among children in different tracks. She cites a high school in Rockford IL in which the math scores of students in high-track courses ranged from the 26th to the 99th percentile on national achievement tests. In the same school, the scores of students assigned to lower tracks ranged between the 1st and 99th percentile. (Note 38) Oakes reports similar score patterns in various subjects throughout most of the middle and high schools in the Rockford and San Jose districts.

In many cases, race and class appear to be better predictors of track placement than any academic measure. For example, Oakes found that in San Jose, white students with average math on national tests scores were three times more likely to be placed in high-track math courses than Latino students with similar scores. The discrepancies for students with higher scores are even more striking: For students scoring between 90 and 99th percentile on national tests, only 56% of Latinos were placed in high-track courses, compared to 93% of whites and 97% of Asians. Similar patterns of discrimination were found at the senior and junior high levels in Rockford.

In a district in Southern California, 88% of white students who scored in the top quartile on the Comprehensive Test of Basic Skills were placed in algebra classes; but only 42% of Latino and 51% of African American students who scored in the top quartile were were placed in algebra. For students who scored in the second quartile, 11% of Latino and 16% of African American students were placed in algebra, compared to 83% of Asian and 53% of white students. (Note 39) Mickelson found similar patterns in the Charlotte-Mecklenburg school district, where white students were far more likely than black students of equal tested ability to be assigned to higher mathematics, laboratory science, and advanced courses in English and history. These results held even when the researchers controlled for students' prior achievement, level of effort, and parents' education. (Note 40)

Oakes also found that the same student might be in one track or another depending on the district or school she attends. Students who might be allocated to a college preparatory track in one school district would, in probability, be assigned to dead-end general or vocational tracks in another (Note 41)

Placement in lower tracks virtually guarantees that students are taught more slowly, exposed to more rudimentary content, and given high grades for work that would, in other settings, be considered unacceptable. For example, Oakes found that students in low-track science and mathematics courses were given more worksheets, tests, and other rote forms of instruction than the average- and high-track students. (Note 42) She also reports that students in high-track classes at a disadvantaged school frequently have less qualified teachers than students in low-track courses at a more advantaged school. Mickelson found that students in lower tracks are more likely to have teachers who lack training in the field they are teaching. (Note 43)

Several authors have documented the consequences of track placement for students' academic success, high school graduation, completion of higher education, and lifelong income chances. Recently, Rose and Betts have shown how valuable exposure to rigorous college preparatory courses, especially advanced mathematics, can be for minority students. (Note 44)

Besides tracking, labeling students with disabilities is another way schools can separate students from higher-level courses. A state-by-state analysis by Parrish found that in 38 states, African American students were more than twice as likely as white students to be identified as mentally retarded. (Note 45) In 29 states, African American students were more than twice as likely to be identified as emotionally disturbed. Nationally, while African American students account for 14.8 percent of the school age population, they comprise 34.3 percent of students identified with mental retardation and 26.4 percent of students identified as emotionally disturbed. (Note 46) Students labeled in these ways are usually separated from regular classes and taught in "resource rooms" in which teachers instruction focuses on low-level skills.

Oswald and colleagues found that the likelihood of a being labeled mentally retarded or emotionally disturbed varies from district to district. Districts with the lowest proportions African American students are the most likely to identify those students as emotionally disturbed. (Note 47) According to Ladner and Hammonds, (Note 48) in predominantly white districts in Texas, nearly 1 in 4 African American students is assigned to special education. (Note 49)

Even more than placement in lower academic tracks, assignment to special education marks students for academic failure. According to the National Longitudinal Transition Study of Special Education African Americans identified as emotionally and behaviorally disturbed had a 66% failure rate in school. The failure rate for whites so labeled was 38%. African American students with EBD were twice as likely to exit school by dropping out (58.2%) as by graduating (27.5%). (Note 50)

We do not claim that lower track placements and assignment to special education is always inappropriate. There must be students who do better in those programs than they would in regular or advanced classrooms. However, as these data show, conventional public education uses low-track placement and disability labels liberally, especially for disadvantaged students. The result is often a kind of segregation more complete, and more consequential from minority students, than segregation based openly on race.

Misallocation of Opportunity-expanding Programs

The fact that students must all be taught to read and do basic arithmetic defines most elementary schools, and limits the degree to which they can differ from one another. Among the public elementary schools in a given district, the most important differences are due to variations in staff quality, or to school culture difference resulting from habits of staff interaction. Beyond those differences some schools get programs that others do not. Not every school gets a special program for gifted and talented students. Many districts offer one or two schools designed on a distinctive model of instruction, like Montessori. Gifted programs and special schools based on brand-name instructional approaches are allocated on a squeaky wheel basis, either to neighborhoods with activist parents or to areas of town where parents are beginning to depart for private or suburban schools. Thus in most districts, such programs and districts are disproportionately available to middle class, usually white, children.

High schools have much more varied programs. Not every school has excellent laboratories, an array of advanced placement courses, or enough qualified teachers of mathematics, science, or languages to allow every student to pursue an advanced college preparatory course. These opportunities are allocated in part by traditional course taking patterns in a school, an approach that sounds reasonable but can create a self-fulfilling prophesy: students in a school where few students formerly took advanced courses lose any opportunity to take such courses. These opportunities are also allocated in response to family and neighborhood pressure, which further favors schools serving middle class students.

This process is not always one way, however. Urban districts facing criticism about low-performing schools in poor neighborhoods sometimes assign reputedly "successful" schools from middle class neighborhoods to these schools. Families in the "nicer" schools often feel deprived in this way, and schools often face difficult adjustments when a principal is pulled out of a smoothly functioning school.

The Baseline Allocation of Opportunity-Expanding Programs

Nationally, both African American and Hispanic children are much less likely to be assigned to gifted programs than students from other groups. According to the Office of Civil Rights (OCR) in 1992 African American students were 57% as likely, and Hispanic students 58% as likely, as children from other groups to be considered gifted. (Note 51)

Economically disadvantaged students are also significantly underrepresented in gifted education. Only 9 percent of students in gifted and talented education programs were in the bottom quartile of family income, while 47 percent of program participants were from the top quartile in family income. (Note 52)

Another measure of minority students' separation from opportunity-expanding programs is their low participation in advanced placement (AP) courses. These are often the most advanced courses offered by high schools, and students who attain high scores on national tests can gain college credit. Nationally, African American and Latino students are far less likely than white and Asian students to take AP courses. Statewide AP data for Texas also fit this pattern. In 1998-99, 10.9 percent of all high school students, but only 4.2% of African American and 7.1% of Hispanic students, took AP courses. However, African American and Hispanics are also less likely than others to score 3 or above on the tests: 31% and 48% compared to 58% of all AP-takers.(Note 53)

To some degree, however, these figures might reflect differences among school districts – especially since minority students cluster in districts that do not offer many or any AP courses for anyone. Within-district data are more telling about the consequences of bureaucratic processes. As Berhholc and colleagues have shown for one district (Wake County North Carolina) African American students make up 24% of the high school population but only 3.5% of students taking AP examinations. (Note 54) The corresponding percentages for Hispanic students are 2.3 and 1.8, and for whites 70 and 78. Of course, AP courses are meant only for well-prepared students, and enrollment differences might reflect the numbers of different groups prepared for these courses. This might explain some of the exclusion of black students, since only 56% of those who took AP courses (compared to 78% of white students) got scores equal to or above 3, usually considered the threshold for college credit. This pattern is reversed, however, for Hispanic students: 87% of those who took AP courses made scores of 3 or above.

Oakes and colleagues had similar findings when comparing low- and high-income neighborhood schools in the Los Angeles Unified School District. Of 12 very large high schools in low-income neighborhoods, only 639 students took AP exams in math and science and only 18%, or 117 students, earned a score of 3 or above. Conversely, 5 high schools in the district's high-income neighborhoods had 890 students take the math and science AP exams, with 71% or 629 students receiving a pass score.

Table 1 summarizes what we have learned about the baseline against which choice programs should be compared. The next section summarizes what little we know about the effects of choice programs.

Table 1
Best Estimates on Incidence of Segregated Placements And Resource Inequities

	Current System **Performance**
Racially Isolated Schools	Schools often exceed district-wide average proportion black or white by 20% or more
Inequitable Allocation of Dollars and Teachers	Most experienced and expensive teachers cluster in "nicest" neighborhoods, per-pupil expenditures unequal

Inequitable Allocation of Opportunity-Expanding Programs	White and middle class children 3 times more likely to enroll in gifted and AP programs
Inequitable Allocation of Opportunity-Limiting Programs	Minority and lower income children 3 times more likely to be enrolled in lower tracks and out-of-class special education

What is Known About Choice Programs

Choice-based programs, whether based on vouchers or school chartering, must confront the same realities that limit the current system: economics, neighborhood segregation, fertility trends, and costs of transportation. Critics and supporters of choice differ on whether it is likely to increase or decrease segregation and inequities in the allocation of dollars, quality teachers, and opportunity-limiting or opportunity-expanding programs.

With respect to segregation, critics of choice fear that it can exacerbate the problem by allowing privileged families to take advantage of their superior access to information to select the best schools; by tolerating admissions processes that let privileged families monopolize access to the most attractive schools; and by allowing the most sought-after schools to hand-pick the easiest-to-educate students.

Defenders of choice programs would respond that these abuses could be eliminated by good program design. Choice programs can promote desegregation in ways conventional public school systems do not – by encouraging out-of-neighborhood school placement, allowing formation of new schools accessible to students in overcrowded schools, and by encouraging expansion or reproduction of oversubscribed schools.

With respect to dollar and human resource inequities, critics fear that choice will lead to heavier financing of schools preferred by privileged families, and concentration of the ablest teachers in schools with the most money and most rewarding students.

Defenders of choice point out that voucher and charter plans all start with transparent allocation of dollars to schools and equality of per-pupil spending. Supply-side choice also constrains schools to live within defined real-dollar budgets, so that no school can afford to hire all the highest-paid teachers. Choice supporters admit, however, that there is nothing to prevent schools with the best reputations hiring the very best teachers or using their funds much more efficiently than other schools.

With respect to opportunity-limiting programs, schools of choice could come under the same pressures as existing public schools, to avoid slowing down faster students by creating lower-track programs for the disadvantaged. Organizations that ran networks of several schools (e.g. like charter school networks or Catholic archdioceses) could also create specialty schools specially targeted to children of different ability levels. Some "special" schools and programs might become unchallenging and low status, and students might be assigned to them on the basis of race or social class.

Defenders of choice argue that competition makes these results unlikely: Schools that create highly differentiated programs will be inefficient and lose out to schools that offer a limited number of focused courses; (Note 55) and families will leave schools that put their children in dead-end courses. There is some favorable evidence about existing schools of choice: charter schools and parochial schools offer more restricted sets of courses than public schools, and parochial schools make sure that disadvantaged students experience mainstream college prep courses. (Note 56) These facts, however, apply to a limited number of schools of choice, most operated by groups with strong commitments to social justice. No one can say for sure whether some schools in a much larger school choice sector might allocate minority-students to opportunity limiting programs.

With respect to opportunity-expanding programs, under any choice scheme, entrepreneurs (charter school operators, nonprofit organizations, for-profit contractors) could choose to locate their schools in areas more accessible to "easy to educate" children. Competition will naturally limit the number of schools that can succeed by this strategy, but poorer neighborhoods could still get more "bare bones" schools. This could happen for two reasons: school providers could decide there is insufficient demand for advanced courses of study in poorer neighborhoods; and organizations running more than one school could try to run lower-cost operations in poorer neighborhoods in order to subsidize the more excellent programs needed to compete in richer neighborhoods.

Defenders of choice argue that school providers have a strong incentive to demonstrate that they can serve the populations that public schools now serve badly. They point to evidence that organizations that manage many schools of choice serve a lower-income and more heavily minority clientele than their surrounding school districts. (Note 57)

Why Evidence on the Effects of Choice is Limited

Empirical evidence is thin on all sides of these arguments. Current voucher and charter school programs are small in scale and many are focused on serving poor and minority children. The results of those programs show that some independently run schools will serve the disadvantaged. But they do not prove that systems of universal choice would have the same benign results.

The evidence is incomplete in another way: current voucher and charter programs do not have the kinds of supply-side effects that universal choice programs are likely to have. Groups that start new schools must now accept less money per pupil than public schools get, and they know that the charter or voucher program on which they rely could be canceled almost at any time. Starting a new school would be a much easier proposition if children came with the full public per-pupil expenditure and if choice programs were stable. Until such a program exists we cannot know how many new schools will arise, or what courses of instruction they will offer, or whom they will serve. (Note 58)

It is important to say why the evidence is so thin. Most choice-oriented policies, including charter school laws and voucher initiatives, are constructed politically. Groups like teachers unions and school administrators associations oppose such policies, but when it is obvious that some forms of choice will be permitted, they focus on limiting their size and scope. (Note 59) By these processes, groups opposing the original voucher program in Alum Rock succeeded on constraining it so that few parents had choices and few new schooling options were created. (Note 60) Today, groups opposing voucher programs work to limit the numbers of families that may choose and the numbers of schools that can be chosen. Opponents also work to limit the amount of money that follows children to schools of choice, often ensuring that charter schools and private schools accepting vouchers receive less money per pupil than is spent in local public school districts. Moreover, teachers' unions and school boards often unite to cushion public schools from the financial impact of losing students. (Note 61)

Taken together, such constraints on choice programs limit what can be learned from them. Limits on who may choose schools can bias choice programs – in some cases toward serving disproportionate numbers of poor or minority children, and in some cases toward excluding poor families that cannot pay extra tuition or provide volunteer services that under-funded schools must require.

Table 2 illustrates the kinds of constraints that have been imposed on choice programs, on both the supply and demand sides. No wonder the evidence about how choice would work in the real world is so limited.

Table 2
Constraints that Reduce the Evidence Value of Choice Programs

Supply Side Constraints	Demand Side Constraints
Rules limiting the numbers of schools of choice that may be created [1, 6]	Limits on the numbers of students (or the percentage of students in a locality) who may choose schools [1, 2, 3, 6]
Rules preventing private groups from operating publicly-funded schools [1, 4]	Rules eliminating former private school students from receiving vouchers [2]
School board refusal to approve more than token numbers of charters [5]	Rules allowing only students with certain characteristics (e.g. poverty or racial minority status) to choose schools [2, 3]
Laws allowing only existing public schools to receive charters [4]	Limits on the neighborhoods from which a family may choose schools [1, 4]
Regulations controlling who may teach in schools, what methods they employ, and how they use time and money [1, 4]	"Legacy" arrangements that give families who live near a school first choice of whether to attend it [1]

Lower per pupil funding for vouchers or for charter schools (relative to district run schools [2, 3, 6]	Rules limiting family choice only to schools that will accept small vouchers (less than public per pupil expenditure) as full tuition [2, 3]

Legend:

1. Alum Rock voucher program (Note 62)
2. State-funded voucher programs in Milwaukee and Cleveland (Note 63)
3. Private voucher programs, e.g., those sponsored by CEO America
4. Weak charter school laws, e.g., Georgia's (Note 64)
5. Charter school laws that do not establish criteria for school board approval of charters
6. Virtually all charter school laws

Conclusion

Until a serious choice experiment is tried – one large and long-lasting enough to gauge supply-side effects as well as families' decisions – we cannot say for sure whether choice would provide worse outcomes than the current system, or how tightly choice must be regulated.

For the time being however, it appears that opponents of choice and defenders of the current public system have inappropriately assigned the burden of proof. Opponents condemn choice because it creates opportunities for alert and aggressive parents to gain the best of everything for their children. They argue that choice is risky and that the existing public education system is a safer and more just alternative. However, as this paper has shown, the existing public education system, which restricts choice by assigning children to schools and limiting the supply of available publicly-funded schools does not accomplish desegregation or give disadvantaged children equitable access to good schools. Public school systems are segregated, particularly in the big cities where poor and minority children are most concentrated. This is so despite decades of serious effort and unwavering declaratory policy in favor of desegregation and equity.

The existing public education system also creates inequities that might not occur under choice: it allows the best-paid teachers to cluster in middle class schools, causing serious within-district inequities in per-pupil spending. It allocates excellent learning opportunities, including advanced placement courses and programs for the gifted, disproportionately to schools serving higher-income children of well-educated parents. It assigns poor and minority students disproportionately to low-track courses, and assigns minority children – particularly African American males – to forms of special education that separate them from regular classes and virtually guarantee that they will drop out before graduating from high school.

Not all these actions on the part of the existing public education system are unambiguously harmful: some children benefit from placements outside the college prep sequence and some children need treatment for emotional disturbance even if that means they miss class. Any system of publicly-funded education, whether based on universal choice or run by a public monopoly, would need some special programs for severely disruptive children or children who need unusual forms of instruction.

Choice programs must not be ruled out because they can lead to some inequities. Every system of allocating opportunities known to man creates some inequities. No matter how opportunities are allocated, parents will seek the best for their own children. Systems should be designed to minimize inequities, and programs should be compared according to the scope and seriousness of inequities they permit.

Choice programs must be carefully designed to prevent segregation, and any program that produced levels of segregation as great as those now prevailing in the public education system should be scrapped or redesigned. (Note 65) Designers of choice programs cannot be expected to eliminate discrimination entirely. But it is fair to demand that they prevent it more effectively than do the bureaucratic processes of conventional public school systems.

Acknowledgement

We are grateful to Jacob Adams for an especially demanding and constructive review of an earlier draft.

Notes

1. Throughout this article we use the term "critics of choice" to refer to scholars and analysts who fear that choice will harm the interests of the poor and disadvantaged. These critics include: Fuller, B. (1996). School Choice: Who Gains, Who Loses? *Issues in Science and Technology*. 12(3) pp. 61-67; and Fuller, B. (1996). Is School Choice Working? *Educational Leadership*. 54(2) pp.37-40.

Concludes that choice may worsen racial separation in schools.

Smith, K. B. & Meier, K. J. (1995). School Choice: Panacea or Pandora's box? *Phi Delta Kappan*. 77(4) pp. 312. Conclude that families choose schools in order to associate with others of the same religion and to avoid racial minorities.

Elmore, R. F. & Fuller, B. (1996). Empirical Research on Education Choice: What are the Implication for Policy-Makers? In Fuller, B., Elmore, R. F., & Orfield, G. (Eds.) *Who Chooses Who Loses?* New York: Teachers College Press. "Increasing educational choice is likely to increase separation of students by race, social class, and cultural background" (p.189). Elmore et. al. argue that regardless of the choice program design, the differences in choosers and non-choosers are such that choice programs will contribute to social stratification, not greater equality.

Wells, A. S. (1998). Charter School Reform in California: Does it meet expectations? *Phi Delta Kappan*. 80(4) pp. 305-312. Argues that charter schools will worsen inequality.

Schneider, M., Marschall, M., Teske, P., & Roch, C. (1998). School Choice and Culture Wars in the Classroom: What Different Parents Seek form Education. *Social Science Quarterly*. 79(3) pp. 489-501.Argue that school choice will increase segregation because parents of different ethnicities and SES status have fundamental differences in their expectations of education for their children.

2. See Fred M. Newmann, Bets Ann Smith, Elaine Allensworth, and Anthony S. Bryk, (2000) *School Instructional Program Coherence: Benefits and Challenges* (Chicago: Consortium on Chicago School Research).

3. Haycock, Kati et. al., (2000) *Achievement in America 2000*, Washington D.C, The Education Trust/

4. See for example, Rose, Heather and Julian R. Betts, (2001) *Math Matters: The Links Between High School Curriculum, College Graduation and Earnings*, San Francisco, The Public Policy Institute of California/

5. Abby Goodnough, (2001) How to Get Your Child the Right Teacher Next Fall

   New York Times Magazine, Sunday May 13.

6. Abby Goodnough, (2001), *op cit.*

7. See Kohn, Alfie,(1998) "Only for My Kid: How Privileged Parents Undermine School Reform", *Phi Delta Kappan*, April 1998, pp. 569-577.

8. Gorard, S., J. Fitz, and C. Taylor, (2001) School Choice Impacts: What Do we Know? *Educational Researcher*, 30, no. 7, October 2001, p. 22.

9. Gary Orfield, (2001) *Schools More Separate: Consequences of a Decade of Resegregation*, Cambridge MA, The Civil Rights project, Harvard University, p.10

10. Throughout this article we will focus on differences in opportunity *within* school districts.

11. See Gary Orfield and John T. Yun, (1999) *Resegregation in American schools*. Cambridge MA, The Civil Rights project, Harvard University.

12. See, for example, Gary Orfield, (2001) *Schools More Separate: Consequences of a Decade of Resegregation*, Cambridge MA, The Civil Rights project, Harvard University, p. 28. Orfield does not try to estimate the growth in segregation due to differential fertility.

13. Orfield and Yun, p. 17.

14. Orfleid 2001, p. 33

15. Orfield 2001, p. 20

16. Orfield 2001, p. 29

17. Orfield 2001 p. 47

18. Orfield 2001, p. 41

19. Michal Kurlaender and John T. Yun, (2000) *Is Diversity a Compelling Educational Interest? Evidence from Metropolitan Louisville*, Cambridge MA, The Civil Rights Project, Harvard University, p.8.

20. Data re-analyzed by the present authors from Stephen Samuel Smith and Roslyn Arlyn Mickelson, (2000) All that Glitters is not Gold: School Reform in Charlotte-Mecklenburg, *Educational Evaluation and Policy Analysis*, vol. 22, no 3, summer, p.101-128

21. Orfield 2001 p. P. 27

22. Parrish, T. (DRAFT 2002). *Disparities in the Identification, Funding, and Provision of Special Education*. Submitted to The Civil Rights Project for The Conference on Minority Issues in Special Education in Public Schools. http://www.law.harvard.edu/civilrights/conferences/SpecEd/parrishpaper2.html

23. Since the early 1970s there has been a research and litigation industry focused on differences in per pupil expenditure among the school districts in a state. Courts have repeatedly found that state policies leading to unequal per-pupil funding violate the equal protection clause of the 14th Amendment to the U.S. Constitution, This industry has largely ignored the dramatic differences in spending and resource allocation *within* school districts. Presumably, the same Constitutional principles could be applied to the inequities identified in this section.

24. Roza, Marguerite and Paul T. Hill, (forthcoming 2004) How Within-District Spending Inequalities Help Some Schools to Fail, in Ravitch, Diane (ed.) *Brookings Papers on Education Policy 2004*, Washington, Brookings Institution Press.

25. Source: Personal communication with Dr. Susan Sclafani, former Houston Deputy Superintendent

26. Education Watch Online: New State and National Achievement Gap Report. The Education Trust. www.edtrust.org

27. See the Education Watch Online website for individual state information.

28. Betts, J. R., Rueben, K. S., & Danenberg, A. (2000). Equal Resources, *Equal Outcomes? The Distribution of School Resources and Student Achievement in California*. Public Policy Institute of California

29. Education Watch Online website

30. Haycock, Kati et. al., (2000) *Achievement in America 2000*, Washington D.C, The Education Trust.

31. Oakes, J. (1990). *Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science*. Santa Monica, CA: Rand.

32. Educational Watch Online

33. In the preceding section on segregation we focused on how students are allocated among schools. This section focuses on how students are allocated to classes and programs within schools.

34. J, Kimbrough and P. T. Hill (1981) *The Aggregate Effects of Federal Education Programs*, Santa Monica CA, RAND.

35. See Bryk, Anthony S., Valerie Lee, and Patrick Holland (1993) *Catholic Schools and the Common Good*, Cambridge MA, Harvard University Press; See also Hill, P.T, G. Foster, and T. Gendler, (1990) *High Schools with Character*, Santa Monica CA, RAND.

36. See Loveless, T., (1999) *The Tracking Wars: State Reform Meets School Policy*, Washington D.C., Brookings.

37. Oakes, J. (1985). *Keeping Track: How Schools Structure Inequality*. New Haven: Yale University Press.

38. Oakes, J. (1995). Two Cities' Tracking and Within-school Segregation. *Teachers College Record.* 96(4) pp. 681-690.

39. The Achievement Council, Inc. Los Angeles, CA. Unpublished. 1991. As cited in Achievement in America 2000, The Education Trust, Inc. http://204.176.179.3 6/dc/edtrust/edstart.cfm

40. Mickelson, R. A. (2001). Subverting Swann: First- and Second-Generation Segregation in the Charlotte-Mecklenburg Schools. *American Education Research Journal.* 38(2) pp.215-252.

41. Oakes, J. & Guiton, G. (1995). Matchmaking: The Dynamics of High School Tracking Decision. *American Educational Research Journal.* 32(1) pp. 3-33.

42. Oakes, J. (1990). *Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science.* Santa Monica, CA: Rand.

43. Mickelson (2001) p. 238

44. Rose, Heather, and Julian Betts (2001) *Math Matters: The Links Between High School Curriculum, College Graduation, and Earnings,* San Francisco,Public Policy Institute of California, 2001

45. Parrish, T. (DRAFT). Disparities in the Identification, Funding, and Provision of Special Education. Submitted to The Civil Rights Project for The Conference on Minority Issues in Special Education in Public Schools. http://www.law.harvard.edu/civilrights/conferences/SpecEd/parrishpa per2.html

46. U. S. Department of Education, Office of Special Education Programs. (2000). *Twenty-second Annual Report to Congress on the Individuals with Disabilities Education Act.* http://www.ed.gov/offices/OSERS/OSEP/OSEP2000AnlRpt/

47. Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic Representation in Special Education: The Influence of School-Related Economic and Demographic Variables. *The Journal of Special Education.* 32(4) pp. 194-206.

48. In studies of district data from Texas and Florida, Lander and Hammons also found that race influences special education rates more than other predictor variable such as poverty, student-teacher ratio, spending per pupil, and teacher salaries. The effect of race is almost double the next highest variable (poverty) and is stronger than the combination of the other three variables in this study. They also present data suggesting that African American and Hispanic students' placement rate in special education is nearly 10% higher in predominately white districts than in predominately minority districts.

49. Ladner, M. & Hammons, C. (2001). Special but Unequal: Race and Special Education. In Finn, Rotherham, & Hokanson (Eds.) *Rethinking Special Education for a New Century.* Fordham Foundation and the Progressive Policy Institute.

50. Valdes, L.A., Williamson, C.L., & Wagner, M.M. (1990). *The National Longitudinal Study of Special Education Students, Statistical Almanac,* Vol. 3: Youth Categorized as Emotionally Disturbed. Menlo park, CA: SRI International. As cited in Osher, D., Woodruff, D., & Sims, A. (DRAFT). Exploring Relationships between Inappropriate and ineffective Special Education Services for African American Children and Youth and the Overrerpresentation in the Juvenile Justice System. www.law.harvard.edu/civilrights/conferences/SpecEd/osherpaper2.html

51. Ford, D. Y. (1998). The Underrepresentation of Minority Students in Gifted Education: Problems and Promises in Recruitment and Retention. *The Journal of Special Education.* 32(1) pp. 4-14.

52. National Education Longitudinal Study (NELS) of 1988 in National Center for Education Statistics, *Urban Schools: The Challenge of Locational Poverty* (Washington, D.C.: U.S. Department of Education, 1996).

53. Texas Education Agency, Office of Policy Planning and Research (2000). *Advanced Placement and International Baccalaureate Examination Results in Texas,* 1998-99.

54. Bernholc, A., Baenen, N., & Howell, R. (2000). *Measuring Up: 1998-99 Advanced Placement Exam Results.* Wake County Public Schools. Evaluation and Research Department.

55. See Hill, P.T. (1999). The supply side of choice. In F. Kemmerer & S. Sugarman (Eds.), *School choice and social controversy.* Washington, D.C.: Brookings Institution Press.

56. See Coleman, James S., and Thomas Hoffer, (1987) *Public and Private High Schools*, New York, Basic Books; See also Coleman, James, Thomas Hoffer, and Sally Kilgore, (1982) *High School Achievement: Public, Catholic, and Private Schools Compared,* New York, Basic Books.

57. With respect to charter schools see U.S. Department of Education, *The State of Charter Schools 2000*, Washington, January 2000, Sec. C p. 2. Nationally, white students make up 48% of the charter school population compared to 58% of the population served by conventional public schools. Charter school student populations are disproportionately white in Arizona, California, Colorado, and Georgia, and disproportionately minority in Florida, Massachusetts, Michigan, Minnesota, New Jersey, North Carolina, Pennsylvania, Texas, and Wisconsin.

58. The British experience with choice shows that large-scale choice programs have much more equitable effects than do small-scale programs, and that results become more equitable the longer a choice program is in place. See Gorard, S., J. Fitz, and C. Taylor, School Choice Impacts: What Do We Know?

59. See Bulman, R.C. & Kirp, D.L. (1999). The Shifting Politics of School Choice. pp.36-67. *In School Choice and Social Controversy: Politics, Policy, and Law*. Eds. Sugarman & Kremerer. Washington, D.C.

60. See Weiner, S.S. & Kellan, K. (1974). *The Politics and Administration of the Voucher Demonstration in Alum Rock, The First Years, 1972-1973.* Santa Monica, CA: RAND.

61. See National Governors' Association (1993). *Strategic Investment: Tough Choices for America's Future*. Washington, D.C.: Brookings Institution Press.

62. See, for example, Henig, J.R. (1994). *Rethinking School Choice: Limits of the Market Metaphor*. Princeton University Press: Princeton, N.J.

63. See Peterson, P.E., Greene, J., & Noyes, C. (1996). School Choice in Milwaukee. *Public Interest*. 125. pp.38-56.

64. Re. charter school laws, see Hill, P.T. and Lake, R.J., (2002) *Charter Schools and Accountability in Public Education*, Washington D.C., Brookings, ch. 4. See also Hassel, B. (1999). *The Charter School Challenge: Avoiding the Pitfalls, Fulfilling the Promise*. Brooking Institution Press: Washington, D.C. See also Center for Education Reform (2001). *Charter School Laws Across the States.* http://edreform.com/charter_schools/laws/

65. Among serious analysts even those most worried about choice admit that ensuring equity is a matter of thoughtful program design. From Cobb and Glass 2000: "The social consequences of choice in education are mediated by the policies under which choice operates. Depending on the degree of public oversight, choice can serve contradictory purposes. Consider two extreme scenarios. Under regulated conditions, choice can correct for severe levels of segregation and ensure the stable integration of schools (e.g., controlled open enrollment plans, magnet programs). Minneapolis, Minnesota and Cambridge, Massachusetts endorse such policies. Conversely, unregulated choice can intensify ethnic stratification by allowing parents to remove their children from integrated schools (e.g., White flight). Arizona's laissez-faire charter legislation appears to fall in this latter group."

References

Bernholc, A., Baenen, N., & Howell, R. (2000). *Measuring up: 1998-99 Advanced Placement Exam results.* Wake County Public Schools. Evaluation and Research Department.

Betts, J. R., Rueben, K. S., & Danenberg, A. (2000). Equal resources, equal outcomes? The distribution of school resources and student achievement in California. Public Policy Institute of California.

Bryk, A. S., Lee, V. & Holland, P. (1993). *Catholic schools and the common good.* Cambridge, MA: Harvard University Press.

Bulman, R.C. & Kirp, D.L. (1999). The shifting politics of school choice. In F. Kemmerer & S. Sugarman (Eds.), *School choice and social controversy.* Washington D.C.: Brookings Institution Press.

Center for Education Reform (2001). *Charter School Laws Across the States.* http://edreform.com/charter_schools/laws/

Cobb, C.D. & Glass, G.V. (1999). Ethnic segregation in Arizona charter schools. *Education Policy Analysis Archives*. 7(1). Retrieved October 1, 2003 from http://epaa.asu.edu/epaa/v7n1/

Coleman, J.S. & Hoffer, T. (1987). *Public and private high schools*. New York, NY: Basic Books.

Coleman, J., Hoffer, T. & Kilgore, S. (1982). *High school achievement: Public, Catholic, and private schools compared*. New York, NY: Basic Books.

Education Watch Online. *New State and National Achievement Gap Report*. The Education Trust. www.edtrust.org

Elmore, R. F. & Fuller, B. (1996). Empirical Research on Education Choice: What are the Implication fro Policy-Makers? In B. Fuller, R.F. Elmore, & G. Orfield, (Eds.) *Who chooses who loses?: Culture, institutions and the unequal effects of school choice*. New York, NY: Teachers College Press.

Ford, D. Y. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *The Journal of Special Education*, *32*(1), 4-14.

Fuller, B. (1996). School Choice: Who Gains, Who Loses? *Issues in Science and Technology*. 12(3), 61-67

Fuller, B. (1996). Is School Choice Working? *Educational Leadership*, *54*(2), 37-40.

Goodnough, A. (2001, May 13). How to get your child the right teacher next fall, *New York Times Magazine*.

Gorard, S., Fitz, J. & Taylor, C. (2001). School choice impacts: What do we know? *Educational Researcher*. *30*(7), 22.

Hassel, B. (1999). *The charter school challenge: Avoiding the pitfalls, fulfilling the promise*. Washington, D.C.: Brookings Institution Press.

Haycock, K. (2000). *Achievement in America 2000*, Washington, D.C.: The Education Trust.

Henig, J.R. (1994). *Rethinking school choice: Limits of the market metaphor*. Princeton, N.J: Princeton University Press.

Hill, P.T. (1999). The supply side of choice. In F. Kemmerer & S. Sugarman (Eds.), *School choice and social controversy*. Washington, D.C.: Brookings Institution Press.

Hill, P.T, Foster, G, and Gendler, T. (1990) *High schools with character*. Santa Monica, CA: RAND.

Hill, P.T. & Lake, R.J., (2002) Charter schools and accountability in public education, Washington, D.C.: Brookings Institution Press.

Hoxby, C. (2000) Does competition among public schools benefit students and taxpayers? *American Economic Review*, *90*(5), 1209-1238.

Kimbrough, J. & Hill, P.T. (1981). *The aggregate effects of federal education programs*, Santa Monica, CA: RAND.

Kohn, A. (1998) Only for my kid: How privileged parents undermine school reform. *Phi Delta Kappan*, *79*(8), 569-577.

Kurlaender, M. & Yun, J.T. (2000). *Is diversity a compelling educational interest? Evidence from metropolitan Louisville*, Cambridge MA, The Civil Rights Project, Harvard University.

Ladner, M. & Hammons, C. (2001). Special but unequal: Race and special education. In C.E. Finn, A.J. Rotherham, & C.R. Hokanson (Eds.) *Rethinking Special Education for a New Century*. Fordham Foundation.

Loveless, T., (1999) *The tracking wars: State reform meets school policy*, Washington, D.C.: Brookings Institution Press.

Mickelson. R. A. (2001). Subverting Swann: First- and second-generation segregation in the Charlotte-

Mecklenburg schools. *American Education Research Journal*, *38*(2), 215-252.

National Governors' Association (1993). *Strategic investment: Tough choices for America's future.* Washington, D.C.

Newmann, F.M., Smith, B.A., Allensworth, E. & Bryk, A.S. (2000). *School instructional program coherenc: Benefits and challenges*. Chicago, IL: Consortium on Chicago School Research.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.

Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science.* Santa Monica, CA: Rand.

Oakes, J. (1995). Two cities' tracking and within-school segregation. *Teachers College Record*, *96*(4), 681-690.

Oakes, J. & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal*, *32*(1), 3-33.

Orfield, G. (2001). *Schools more separate: Consequences of a decade of resegregation,* Cambridge, MA, The Civil Rights project, Harvard University.

Orfield, G. & Yun, J.T. (1999). *Resegregation in American schools.* Cambridge, MA, The Civil Rights project, Harvard University.

Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. *The Journal of Special Education*, *32*(4), 194-206.

Parrish, T. (DRAFT). *Disparities in the identification, funding, and provision of special education*. Submitted to The Civil Rights Project for The Conference on Minority Issues in Special Education in Public Schools. http://www.law.harvard.edu/civilrights/conferences/SpecEd/parrishpa per2.html

Peterson, P.E., Greene, J., & Noyes, C. (1996). School choice in Milwaukee. *Public Interest*, *125*, 38-56.

Rose, H. & Betts, J. (2001). San Francisco, *Math matters: The links between high school curriculum, college graduation, and earnings.* San Francisco, CA: Public Policy Institute of California

Roza, Marguerite and Paul T. Hill (forthcoming 2004) How Within-District Spending Inequalities Help Some Schools to Fail, in Ravitch, Diane (ed.) *Brookings Papers on Education Policy 2004,* Washington, Brookings Press

Schneider, M., Marschall, M., Teske, P., & Roch, C. (1998). School Choice and Culture Wars in the Classroom: What Different Parents Seek form Education. *Social Science Quarterly*, *79*(3), 489-501.

Smith, K. B. & Meier, K. J. (1995). School choice: Panacea or Pandora's box? *Phi Delta Kappan*, *77*(4), 312-316.

Smith, S.S. & Mickelson, R.A. (2000). All that glitters is not gold: School reform in Charlotte-Mecklenburg, *Educational Evaluation and Policy Analysis*, *22*(3).101-128.

Texas Education Agency, Office of Policy Planning and Research (2000). *Advanced Placement and International Baccalaureate examination results in Texas*, 1998-99.

The Achievement Council, Inc. Los Angeles, CA. Unpublished. 1991. As cited in Achievement in America 2000, The Education Trust, Inc.http://204.176.179.36/dc/edtrust/edstart.cfm

U.S. Department of Education (1996). *Urban Schools: The Challenge of Locational Poverty*. National Education Longitudinal Study (NELS) of 1988. National Center for Education Statistics.

U.S. Department of Education (2000). *The State of Charter Schools 2000*, Washington, D.C.

U. S. Department of Education, Office of Special Education Programs. (2000). Twenty-second annual report to congress on the Individuals with Disabilities Education Act.

http://www.ed.gov/offices/OSERS/OSEP/OSEP2000AnlRpt/

Valdes, L.A., Williamson, C.L., & Wagner, M.M. (1990). The national longitudinal study of special education students. *Statistical Almanac, Vol. 3: Youth categorized as emotionally disturbed.* Menlo Park, CA: SRI International. As cited in Osher, D., Woodruff, D., & Sims, A. (DRAFT). Exploring relationships between inappropriate and ineffective special education services for African American children and youth and the overrerpresentation in the juvenile justice system.
www.law.harvard.edu/civilrights/conferences/SpecEd/osherpaper2.html

Weiner, S.S. & Kellan, K. (1974). *The politics and administration of the voucher demonstration in Alum Rock: The first years, 1972-1973.* Santa Monica, CA: RAND.

Wells, A. S. (1998). Charter school reform in California: Does it meet expectations? *Phi Delta Kappan, 80* (4), 305-312.

About the Authors

Paul T. Hill
Research Professor
Daniel J. Evans School of Public Affairs
Director, Center on Reinventing Public Education
University of Washington, Box 353055
Seattle, Washington 98195-3055
Tel.: 206-685-2214
Homepage: http://www.crpe.org

Paul Hill is director of the Center on Reinventing Public Education, which studies alternative governance systems for public elementary and secondary education. He is also a non-resident Senior Fellow of the Brookings Institution. For Brookings he is now leading a national working commission on choice in K-12 education, which will issue its report in late 2003. Before joining the University of Washington faculty, Paul Hill worked for 17 years as a Senior Social Scientist in RAND's Washington Office where his research focused on the reform of elementary and secondary education. He contributed to studies of defense research, development and acquisition policy. While at RAND he served as Director of Washington Operations (1981-87) and Director of the Education and Human Resources program (1979-80). While a government employee he directed the National Institute of Education's Compensatory Education Study (a congressionally-mandated assessment of federal aid to elementary and secondary education) and conducted research on housing and education for O.E.O.

Kacey Guin
Research Assistant
Center on Reinventing Public Education
University of Washington, Box 353055
Seattle, Washington 98195-3055

Kacey Guin is a research coordinator for the Center on Reinventing Public Education, in the Daniel J. Evans School of Public Affairs at the University of Washington. She has co-authored book chapters on the achievement gap and the assessment of school choice programs. Most recently, she completed a study on the impact of high teacher turnover rates on the organizational functioning of schools.

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple University of Wisconsin	David C. Berliner Arizona State University

Greg Camilli
Rutgers University

Linda Darling-Hammond
Stanford University

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Gustavo E. Fischman
Arizona State Univeristy

Richard Garlikov
Birmingham, Alabama

Thomas F. Green
Syracuse University

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Ontario Institute of Technology

Patricia Fey Jarvis
Seattle, Washington

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Les McLean
University of Toronto

Heinrich Mintrop
University of California, Los Angeles

Michele Moses
Arizona State University

Gary Orfield
Harvard University

Anthony G. Rud Jr.
Purdue University

Jay Paredes Scribner
University of Missouri

Michael Scriven
University of Auckland

Lorrie A. Shepard
University of Colorado, Boulder

Robert E. Stake
University of Illinois—UC

Kevin Welner
University of Colorado, Boulder

Terrence G. Wiley
Arizona State University

John Willinsky
University of British Columbia

EPAA Spanish and Portuguese Language Editorial Board

Associate Editors for Spanish & Portuguese

Gustavo E. Fischman
Arizona State University
fischman@asu.edu

Pablo Gentili
Laboratório de Políticas Públicas
Universidade do Estado do Rio de Janeiro
pablo@lpp-uerj.net

Founding Associate Editor for Spanish Language (1998-2003)
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

María Beatriz Luce(Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga

humberto@servidor.unam.mx

Daniel Schugurensky(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for Resesarch–Brazil (AIRBrasil)
simon@sman.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

home abstracts complete editors submit comment notices search

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (1/2003)